

THE
PHILOSOPHICAL MAGAZINE

EDITOR

PROFESSOR N. F. MOTT, M.A., F.R.S.

EDITORIAL BOARD

SIR LAWRENCE BRAGG, O.B.E., M.C., M.A., D.Sc., F.R.S.

ALLAN FERGUSON, M.A., D.Sc.

SIR GEORGE THOMSON, M.A., D.Sc., F.R.S.

PROFESSOR A. M. TYNDALL, D.Sc., F.R.S.

VOL. XL.—SEVENTH SERIES

JANUARY—DECEMBER 1949

LONDON:

TAYLOR & FRANCIS, LTD., RED LION COURT, FLEET STREET, E.C. 4.

61/7-27-1

“Meditationis est perscrutari occulta; contemplationis est admirari perspicua....Admiratio generat quæstionem, quæstio investigationem, investigatio inventionem.”—*Hugo de S. Victore.*

——“Cur spirent venti, cur terra dehiscat,
Cur mare turgescat, pelago cur tantus amaror,
Cur caput obscura Phœbus furrugine condât,
Quid toties diros cogat flagrare cometas,
Quid pariat nubes, veniant cur fulmina cœlo,
Quo micet igne Iris, superos quis conciat orbes
Tam vario motu.”

J. B. Pinelli ad Mazonium.



CONTENTS OF VOL. XL. (Nos. 300-311)

(SEVENTH SERIES)

NUMBER 300—JANUARY

	Page
I. The Properties of certain Strained Hydrocarbons. By C. A. COULSON, Physics Department, King's College, London, and W. E. MOFFITT, Physical Chemistry Laboratory, Oxford	1
II. Investigation of Soft Radiations by Proportional Counters.—I. By S. C. CURRAN, J. ANGUS, and A. L. COCKROFT, Department of Natural Philosophy, University of Glasgow	36
III. Investigation of Soft Radiations.—II. The Beta Spectrum of Tritium. By S. C. CURRAN, J. ANGUS, and A. L. COCKROFT, Department of Natural Philosophy, University of Glasgow	53
IV. A Contribution to the Theory of Liquid Helium II. By N. F. MOTT, F.R.S., H. H. WILLS Physics Laboratory, University of Bristol	61
V. The Coefficients of Expansion of some Solid Solutions in Aluminium. By W. HUME-ROTHERY, F.R.S., and T. H. BOULTBEE, Inorganic Chemistry Laboratory, University Museum, Oxford	71
VI. An Application of the Theory of Quasi-Properties to the Treatment of Anomalous Strain-Stress Relations. By G. W. SCOTT BLAIR, M.A., D.Sc., F.R.I.C., F.Inst.P., and J. E. CAFFYN, B.Sc., A.Inst.P., National Institute for Research in Dairying, University of Reading	80
VII. On the Dimensions of Physical Magnitudes (Seventh Paper: A Paradox in Dimensional Theory). By HERBERT DINGLE, A.R.C.S., D.I.C., D.Sc.	94
VIII. On the Theory of Ferromagnetic Resonance. By D. POLDER, H. H. WILLS Physical Laboratory, University of Bristol (now at Philips Research Laboratories, Eindhoven)	99
IX. Scattering of X-Rays by the L Electrons of Boron. By KESSAR ALEXOPOULOS and HELEN SCOULUDI, Physics Laboratory, University of Athens	115
X. Notices of New Books and Periodicals received:—	
Prof. ENRICO CASTELLI's <i>Existentialisme Théologique</i>	123
Tables of Bessel Functions of Fractional Order.—Vol. I.	124

NUMBER 301—FEBRUARY

XI. The Magnetic Field of Massive Rotating Bodies. By P. M. S. BLACKETT, The University of Manchester	125
XII. Canonical Transformations of the Hamiltonian in Meson Field Theory. By K. J. LE COUTEUR and L. ROSENFELD, University of Manchester	151
XIII. Anomalous Dielectric Properties of Polycrystalline Titanates of the Perovskite Type. By J. R. PARTINGTON, G. V. PLANER and I. I. BOSWELL	157
XIV. On the Oxidation of Metals at Low Temperatures and the Influence of Light. By N. CABRERA, H. H. WILLS Physical Laboratory, University of Bristol	175
XV. The Application of Mellin Transforms to the Summation of Slowly Convergent Series. By G. G. MACFARLANE, Dr.-Ing., B.Sc., Telecommunications Research Establishment, Ministry of Supply, Malvern	188

	Page
XVI. The Brillouin Zones for the CO_2Al_9 and NiAl_3 Structures. By G. V. RAYNOR and M. B. WALDRON, Department of Metallurgy, The University, Edgbaston	198
XVII. Electron Energies resulting from an Electric Field in a Highly Ionized Gas. By R. G. GIOVANELLI, M.Sc.	206
XVIII. The Flow behind a Stationary Shock. By M. J. LIGHTHILL, Department of Mathematics, University of Manchester	214
XIX. Note on a new Form of the Solution of Reynolds' Equation for Michell Rectangular and Sector-shaped Pads. By Mrs. W. L. WOOD, M.A.	220
XX. A Note on Minimum Integrals in Field Theory. By C. W. KILMISTER and B. H. CHIRGWIN	226
XXI. The Focusing of β -rays in a Prolate Spheroidal Magnetic Field. By H. O. W. RICHARDSON, Department of Natural Philosophy, The University, Edinburgh	233
XXII. Notices of New Books and Periodicals received :—	
Mr. J. W. ARCHBOLD's An Introduction to the Algebraic Geometry of a Plane	245
Rapports et Discussions sur les Isotopes.....	245
Prof. E. A. MILNE's Vectorial Mechanics.....	246
Prof. R. W. JAMES' The Crystalline State.—Volume II. The Optical Principles of the Diffraction of X-Rays.....	246
Mr. T. VOGEL's The Vibrations of elastic Systems in a Sound-field	247
Ionospheric Radio Propagation	247

NUMBER 302—MARCH

XXIII. The Properties of Silver Halides containing Traces of Silver Sulphide. By J. W. MITCHELL, D.Phil., H. H. Wills Physical Laboratory, University of Bristol	249
XXIV. Tank Model for Magnetic Problems of Axial Symmetry. By R. E. PETERLS and T. H. R. SKYRME, Department of Mathematical Physics, The University, Birmingham, 15.....	269
XXV. Variation with Oxygen Pressure of the Thermoelectric Power of Cadmium Oxide. By C. A. HOGARTH, Ph.D., and J. P. ANDREWS, D.Sc., Queen Mary College	273
XXVI. The Electrical Conductivity of Bismuth Fibres.—I. Magneto Resistance and the Crystalline Structure. By B. DONOVAN and G. K. T. CONN, Department of Physics, The University of Sheffield.....	283
XXVII. Laplace Transform Solution of Two-medium Neutron Ageing Problem. By R. BELLMAN, R. E. MARSHAK and G. M. WING, University of California, Los Alamos Scientific Laboratory, Santa Fe, New Mexico....	297
XXVIII. Note on the Effect of Impurities and Cold Work on the Thermoelectric Power of Aluminium. By J. K. GALT, H. H. Wills Physical Laboratory, University of Bristol	309
XXIX. Distribution of Reaction Times for Turbulent Flow in Cylindrical Reactors. By R. C. L. BOSWORTH, Ph.D., D.Sc.	314
XXX. Change of Electrical Resistance of Alloys during Ageing. By Z. MATYÁŠ, The First Physical Institute of University of Technical Sciences in Prague	324
XXXI. The Flow of Viscous Fluids round Plane Obstacles. By R. ROSCOE, Ph.D., Physics Department, King's College, Newcastle-upon-Tyne	338
XXXII. Tables of a Function and Related Functions. By K. MITCHELL, Ph.D., A.F.R.Ae.S., King's College, Newcastle-upon-Tyne	351

NUMBER 303—APRIL

Page

XXXIII. Ionization and Charge Exchange by Fast Ions of Hydrogen and Helium. By J. P. KEENE, Cavendish Laboratory, Cambridge	369
XXXIV. Notes on the Molecular Orbital Treatment of the Hydrogen Molecule. By Prof. C. A. COULSON and Miss I. FISCHER, Wheatstone Physics Department, King's College, London.....	386
XXXV. Angular Distribution, Frequency and Absorption of Slow Single Cosmic Ray Protons. By S. LATTIMORE, Imperial College, London.....	394
XXXVI. The Primary Solid Solution of Silver in Aluminium. By Dr. G. V. RAYNOR and Dr. D. W. WAKEMAN, Department of Metallurgy, University of Birmingham	404
XXXVII. Channel Section Waveguide Radiator. By A. L. CULLEN, B.Sc., Electrical Engineering Department, University College, London ..	417
XXXVIII. Regraduation in Spherically Symmetric Space-times of General Relativity. By T. J. WILLMORE, M.Sc., Ph.D., Durham Colleges.	428
XXXIX. Theory of the High Pressure Mercury Vapour and Cadmium Vapour Discharges. By V. J. FRANCIS, B.Sc., M.I.E.E., A.R.C.S., F.Inst.P.	435
XL. A Note on the Theory of Communication. By J. D. WESTON, Sheffield University	449
XLI. Stresses in Elastic Cylinders in Contact along a Generatrix (including the effect of tangential friction). By EWEN M'EWEN, University of Durham	454
XLII. Determination of the Electron Energy Distribution in Gases from Townsend's Ionization Coefficient. By H. D. DEAS and K. G. EMELEUS, Queen's University, Belfast	460
XLIII. On the Mechanical Forces in Dielectrics. By W. B. SMITH-WHITE, University of Sydney	466
XLIV. Notices of New Books and Periodicals received :—	
Dr. N. ERNEST DORSEY's <i>The Freezing of Supercooled Water</i> ..	479
<i>Science and Technology in China</i>	479
Sir WILLIAM CECIL DAMPLER's <i>A History of Science, and its Relations with Philosophy and Religion</i>	479
Prof. A. D. RITCHIE's <i>Essays in Philosophy and Other Pieces</i> ..	480
<i>Journal of the Physical Society of Japan</i>	480
<i>The Collected Works of J. Willard Gibbs</i>	480

NUMBER 304—MAY

XLV. Magneto-thermal Effects in Ferromagnetics. By E. C. STONER, F.R.S., and P. RHODES, Physics Department, University of Leeds	481
XLVI. Investigation of Soft Radiations by Proportional Counters.—	
III. The Beta Spectrum of Carbon 14. By J. ANGUS, A. L. COCKROFT, and S. C. CURRAN, Department of Natural Philosophy, The University of Glasgow	522
XLVII. Origin of Cosmic Ray Stars. By J. B. HARDING, Imperial College, London	530
XLVIII. Characteristics of the Crater of the High Intensity Carbon Arc. By V. J. FRANCIS, B.Sc., F.Inst.P., M.I.E.E., F. S. HAWKINS, Ph.D., A.R.I.C., and A. H. WILLOUGHBY, B.Sc.(Eng.).....	546
XLIX. Transit-time Deterioration of Space-charge Reduction of Shot Effect. By D. K. C. MACDONALD, Clarendon Laboratory, Oxford	561
L. The Efficiency of the Selenium Barrier-Photocell when used as a Converter of Light into Electrical Energy. By E. BILLIG, Dr.tech., and K. W. PLESSNER, M.Sc., Ph.D., Research Laboratory, Associated Electrical Industries, Aldermaston.....	568

	Page
LI. The Zero-Point Energy of a System of Particles. By R. B. DINGLE, Royal Society, Mond Laboratory Cambridge	573
LII. New Methods for the Numerical Solution of Algebraic Equations. By Professor A. PORTER, M.Sc., Ph.D., and C. MACK, M.A.	578
LIII. Heavy Splinters in Cosmic Ray Stars. By Dr. A. BONETTI and Miss C. DILWORTH, Centre de Physique Nucleaire de Bruxelles. (Plates I. & II.)	585
LIV. Notices of New Books and Periodicals received:—	
Mr. L. MARTON's Advances in Electronics	588
Dr. G. ZENER's Elasticity and Anelasticity of Metals	588
Dr. KATHLEEN LONSDALE's Crystals and X-Rays	588
Dr. L. J. COMRIE's Chamber's Six-figure Mathematical Tables..	588

NUMBER 305—JUNE

LV. The Production of Cosmic Ray Stars. By Sir GEORGE THOMSON, Imperial College, South Kensington.	589
LVI. Mechanism of π -meson Disintegrations. By D. H. PERKINS, Ph.D. (Plates III. & IV.)	601
LVII. The Diffraction of Radio Waves from Meteor Trails and the Measurement of Meteor Velocities. By J. G. DAVIES and C. D. ELLYETT, Physical Laboratory, University of Manchester. (Plate V.)	614
LVIII. The Derivation of Statistical Expressions from Gibbs' Canonical Ensemble. By F. ANSBACHER and W. EHRENBERG, Birkbeck College, University of London	626
LIX. Investigation of Soft Radiations by Proportional Counter.—IV. The Beta Spectrum of Ni ⁶³ . By H. W. WILSON and S. C. CURRAN, Department of Natural Philosophy of the University of Glasgow	631
LX. On the Motion of a Vector Meson in a Homogeneous Magnetic Field. By N. SYMONDS, Dublin Institute for Advanced Studies.	636
LXI. On Sommerfeld's "Radiation Condition." By F. V. ATKINSON, M.A., D.Phil.	645
LXII. On the use of Canonical Transformations for Collision Problems. By Professor W. HEITLER, F.R.S., and Dr. S. T. MA, Dublin Institute for Advanced Studies	651
LXIII. Lattice Defects in Silver Halide Crystals. By J. W. MITCHELL, D.Phil., H. H. Wills Physical Laboratory, University of Bristol.	667
LXIV. An Improved Schmidt Plate. By DOROTHY B. G. HAWKINS, Ph.D., H. H. Wills Physical Laboratory, University of Bristol	670
LXV. Notices of New Books and Periodicals received:—	
J. VON NEUMANN's Les Fondements Mathematiques de la Mécanique Quantique	679
Messrs. S. BHAGAVANTAM and T. VENKATARAYUDU's Theory of Groups and its Application to Physical Problems.	681
Profs. R. COURANT and K. O. FRIEDRICHS's Supersonic Flow and Shock Waves	681
Mr. H. YUKAWA's Progress of Theoretical Physics	682
Prof. W. HEISENBERG's Two Lectures	683
Prof. MAX BORN's Natural Philosophy of Cause and Chance ..	683
Messrs. B. FINZI and M. PASTORI's Calcolo Tensoriale e Applicazioni	684
Dr. H. D. ANTHONY's Science and its Background	684
Mr. JOHN J. TOOHEY's An Elementary Handbook of Logic....	684

NUMBER 306—JULY

	Page
LXVI. Capture of Negative Mesons. By R. HUBY, Department of Theoretical Physics, University of Liverpool.....	685
LXVII. On Kramers' Formula for the Frequency Distribution of the Radiation Emitted by an Electron in a Parabolic Orbit. By K. C. WESTFOLD, Division of Radiophysics, Council for Scientific and Industrial Research, Australia.....	698
LXVIII. Interchange Interaction and Collective Electron Ferromagnetism. By E. P. WOHLFARTH, Ph.D., Department of Physics, Leeds University	703
LXIX. On the Interaction between Mesons. By H. Y. Tzu, University of Manchester.....	717
LXX. Lattice Theory of Dielectric and Piezoelectric Constants in Crystals. By KUN HUANG, Department of Theoretical Physics, University of Liverpool	733
LXXI. A Note on the Transient Response of an Oscillatory Circuit with Recurrent Discharge. By A. M. HARDIE, M.A., B.Sc., A.M.I.E.E., Metropolitan-Vickers Electrical Co. Ltd., Manchester.....	748
LXXII. The Loss of Energy of Slow Negative Mesons in Matter. By R. L. ROSENBERG, Natal University College, South Africa.....	759
LXXIII. Thermal Stresses in Turbine Blades. By M. J. LIGHTHILL (Manchester) and F. J. BRADSHAW (Farnborough).....	770
LXXIV. Notices of New Books and Periodicals received :—	
Dr. A. C. MERRINGTON's Vicometry.....	781
Mr. L. HARTSHORN's Radio-Frequency Heating	781
Mr. ALFRED TARSKI's Cardinal Algebras	781
Mr. COLIN CHERRY's Pulses and Transients in Communication Networks	781
Mr. G. H. HARDY's Divergent Series	782
Mr. J. L. COOLIDGE's The Mathematics of Great Amateurs	782
Anuario del Observatorio Astronomico de Madrid.....	782
Prof. EDWARD BAUER's L'Electromagnetisme, Hier et Aujourd'hui	782

NUMBER 307—AUGUST

LXXV. The Numerical Reduction of Non-Singular Matrix Pencils. By J. M. HAMMERSLEY, Lectureship in the Design and Analysis of Scientific Experiment, University of Oxford.....	783
LXXVI. The Emission of Short-range Alpha Particles from Light Elements under Proton Bombardment.—I. Experimental Method and the Reaction $^{10}\text{B}(\alpha)^7\text{Be}$. By W. E. BURCHAM and JOAN M. FREEMAN..	807
LXXVII. The Characteristics of Radio-Frequency Radiation in an Ionized Gas, with Applications to the Transfer of Radiation in the Solar Atmosphere. By S. F. SMERD and K. C. WESTFOLD, Radiophysics Laboratory, Council for Scientific and Industrial Research, Commonwealth of Australia	831
LXXVIII. Frictional Relaxation Oscillations. By B. R. DUDLEY and Professor H. W. SWIFT	849
LXXIX. Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.—Part I. Observations with Photographic Plates exposed at an altitude of 11,000 feet. By Miss R. H. BROWN, U. CAMERINI, P. H. FOWLER, H. HEITLER, D. T. KING and C. F. POWELL, The H. H. Wills Physical Laboratory, University of Bristol, (Plates VI.—XIII.) ..	862

LXXX. Notices of New Books and Periodicals received :—

Mr. L. L. WHYTE's <i>The Unitary Principle in Physics and Biology</i>	882
Prof. JUSTI's <i>Leitfähigkeit und Leitungsmechanismus fester Stoffe</i>	882
Prof. CULLWICK's <i>The Fundamentals of Electromagnetism</i> ..	882

NUMBER 308—SEPTEMBER

LXXXI. <i>Stresses and Strains in Tube-Drawing.</i> By H. W. SWIFT ..	883
LXXXII. <i>Edge Dislocations in Anisotropic Materials.</i> By J. D. ESHELBY, H. H. Wills Physical Laboratory, University of Bristol	903
LXXXIII. <i>On the Problem of a Notched Plate of an Aeolotropic Material.</i> By H. OKUBO, Institute of High Speed Mechanics, Tōhoku University, Sendai, Japan	913
LXXXIV. <i>Distance Correlations in an Ideal Fermi-Dirac Gas.</i> By P. L. BHATNAGAR and K. S. SINGWI, University of Delhi, Delhi, India ..	917
LXXXV. <i>The Calculation of the Optimum Parameters for a Following System.</i> By C. MACK, M.A.	922
LXXXVI. <i>Investigation of Soft Radiation by Proportional Counters.—V. Use as a Detector of Ultra-violet Quanta and Analysis of the Gas Multiplication Process.</i> By S. C. CURRAN, A. L. COCKROFT, and J. ANGUS, Department of Natural Philosophy, University of Glasgow.	929
LXXXVII. <i>Non-symmetric Stress-Energy-Momentum Tensor and Spin-Density.</i> By A. PAPAPETROU, I.C.I. Research Fellow, University of Manchester	937
LXXXVIII. <i>The Determination of the Charge of Heavy Particles emitted during the Explosive Disintegration of Nuclei.</i> By S. O. C. SØRENSEN, The H. H. Wills Physical Laboratory, University of Bristol. (Plates XIV.–XVI.)	947
LXXXIX. <i>The Energy Distribution of Slow Positive Ion Beams.</i> By G. D. YARNOLD and H. C. BOLTON, University of Nottingham	956
XC. <i>Notices of New Books and Periodicals received :—</i>	
<i>Surface Chemistry (Various Authors)</i>	968
Mr. G. S. RUSHBROOKE's <i>Introduction to Statistical Mechanics</i> ..	968
Sir EDMUND WHITTAKER's <i>From Euclid to Eddington</i>	969
Messrs. W. MAGNUS and F. OBERHETTINGER's <i>Formulæ and Theorems for the Special Functions of Mathematical Physics</i> ..	969

NUMBER 309—OCTOBER

XCI. <i>Plastic Distortion of Non-uniform Sheets.</i> By R. HILL, M.A., Ph.D.	971
XCII. <i>On the Theory of the Magnetic Double Refraction in Non-polar Liquids.</i> By OLLE SNELLMAN, Fil.dr., Institute of Physical Chemistry, Upsala, Sweden	983
XCIII. <i>The Magnetic Double Refraction of Liquid Mixtures.—I. Benzene in Various Solvents.</i> By EDWARD J. BURGE and OLLE SNELLMAN, Fil.dr., The Institute of Physical Chemistry, University of Upsala, Sweden	994
XCIV. <i>Investigation of Soft Radiations by Proportional Counters.—VI. The Beta Spectrum of Sulphur 35.</i> By A. L. COCKROFT and G. M. INSCH, Department of Natural Philosophy, Glasgow University	1014

XCV. Symmetry Changes in Barium Titanate at Low Temperatures and their Relation to its Ferroelectric Properties. By H. F. KAY and P. VOUSDEN, H. H. Wills Physical Laboratory, University of Bristol....	1919
XCVI. Theory of Barium Titanate.—Part I. By A. F. DEVONSHIRE, H. H. Wills Physical Laboratory, University of Bristol.....	1040
XCVII. Theory of the High Pressure Helium Discharge. By V. J. FRANCIS, B.Sc., F.Inst.P., M.I.E.E., A.R.C.S.....	1063
XCVIII. Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.—Part II. Observations at High Altitudes by means of Free Balloons. By U. CAMERINI, T. COOR, J. H. DAVIES, P. H. FOWLER, W. O. LOCK, H. MUIRHEAD and N. TOBIN, The H. H. Wills Physical Laboratory, University of Bristol. (Plates XVII.—XIX.).....	1073
XCIX. Experiments with Nuclear-Track Emulsions Sensitive at Minimum Ionizing Power. By A. C. COATES, A.R.C.S., and R. H. HERZ, D.Phil., Research Laboratories, Kodak Ltd., Harrow, Middlesex. (Plates XX.—XXIII.).....	1088
C. Notices of New Books and Periodicals received :—	
Dr. G. W. SCOTT BLAIR's A Survey of General and Applied Rheology	1093

NUMBER 310—NOVEMBER

CI. Magnetic Properties of Nickel-Cobalt and Related Alloys. By E. P. WOHLFARTH, Ph.D., Department of Physics, Leeds University....	1095
CII. Vacuum Polarization in the Positron Theory. By S. T. MA, Dublin Institute for Advanced Studies, Dublin	1112
CIII. Study of the Wave-forms of Atmosphericics. By S. R. KHASTGIR, D.Sc.(Edin.), F.N.I., and R. ROY, M.Sc.	1129
CIV. Electrodynamics in a Rotating Frame of Reference. By M. G. TROCHERIS	1143
CV. Internal Pair Creation in Na^{24} . By E. R. RAE, M.A., Department of Natural Philosophy, University of Glasgow.....	1155
CVI. Scattering of Pseudoscalar Charged Mesons by Nucleons.—I. By E. CORINALDESI and G. FIELD, Dublin Institute for Advanced Studies.	1159
CVII. Notes on the Validity and Application of the Method of Molecular Orbitals. By C. A. COULSON and H. C. LONGUET-HIGGINS.....	1172
CVIII. Notices of New Books and Periodicals received :—	
Mr. G. WENTZEL's The Quantum Theory of Fields	1177
Dr. L. R. G. TRELOAR's The Physics of Rubber Elasticity..	1177
The Physical Society Reports on Progress in Physics	1178
Mr. G. F. J. GARLICK's Luminescent Materials	1178

NUMBER 311—DECEMBER

CIX. A Technique for Rendering Approximate Solutions to Physical Problems Uniformly Valid. By M. J. LIGHTHILL	1179
CX. The Shock Strength in Supersonic "Conical Fields." By M. J. LIGHTHILL	1202
CXI. A Note on the Identity of Thermal Noise and Shot Noise. By B. MELTZER	1224
CXII. On the Theory of Strength of Quasi-Isotropic Solids. By R. FÜRTH, Birkbeck College, University of London	1227

CXIII. The Magnetic Double Refraction of Liquid Mixtures.—II. Toluene, and Carbon Disulphide, in Various Solvents, and Relation of MDR to Refractive Index. By EDWARD J. BURGE, B.Sc. and OLLE SNELLMAN, Fil.dr., Institute of Physical Chemistry, Upsala, Sweden....	1233
CXIV. On the so-called "Clock-paradox" of Special Relativity. By Professor E. A. MILNE, F.R.S. and G. J. WHITROW	1244
CXV. The Decay of μ -Mesons. By J. H. DAVIES, W. O. LOCK and H. MUIRHEAD, The H.H. Wills Physical Laboratory, University of Bristol	1250
CXVI. The Interpretation of X-Ray Absorption Spectra of Solids. By Y. CAUCHOIS and N. F. MOTT. (Plates XXIV. & XXV.)	1260
CXVII. Notices of New Books and Periodicals received:—	
Mr. J. H. AWBERY's A Textbook on Heat.....	1269
Mr. W. J. H. SPROTT's General Psychology	1270
Mr. ANTONIO FERRI's Elements of Aerodynamics of Supersonic Flows	1270
Mr. D. R. HARTREE's Calculating Instruments and Machines.	1270
Index to Volume	1271

PLATES

- I. & II. Illustrative of Dr. A. BONETTI and Miss C. DILWORTH's Paper on the Heavy Splinters in Cosmic Ray Stars.
- III. & IV. Illustrative of Dr. D. H. PERKINS's Paper on the Mechanism of π -Meson Disintegrations.
- V. Illustrative of Messrs. J. G. DAVIES and C. D. ELLYETT's Paper on the Diffraction of Radio Waves from Meteor Trails and the Measurement of Meteor Velocities.
- VI.—XIII. Illustrative of Miss R. H. BROWN, U. CAMERINI, P. H. FOWLER, H. HEITLER, D. T. KING and C. F. POWELL's Paper on Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.—Part I. Observations with Photographic Plates exposed at an altitude of 11,000 feet.
- XIV.—XVI. Illustrative of Mr. S. O. C. SÖRENSEN's Paper on the Determination of the Charge of Heavy Particles emitted during the Explosive Disintegration of Nuclei.
- XVII.—XIX. Illustrative of Messrs. U. CAMERINI, T. COOR, J. H. DAVIES, P. H. FOWLER, W. O. LOCK, H. MUIRHEAD and N. TOBIN's Paper on Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.—Part II. Observations at High Altitudes by means of Free Balloons.
- XX.—XXIII. Illustrative of Mr. A. C. COATES and Dr. R. H. HERZ's Paper on Experiments with Nuclear-Track Emulsions Sensitive at Minimum Ionizing Power.
- XXIV. & XXV. Illustrative of Miss Y. CAUCHOIS and Prof. N. F. MOTT's Paper on the Interpretation of X-Ray Absorption of Spectra of Solids.

THE PHILOSOPHICAL MAGAZINE

A JOURNAL OF THEORETICAL EXPERIMENTAL
AND APPLIED PHYSICS

First published in 1798

[SEVENTH SERIES—VOL. 40]

I. The Properties of certain Strained Hydrocarbons.

By C. A. COULSON, Physics Department, King's College, London, and
W. E. MOFFITT, Physical Chemistry Laboratory, Oxford*.

[Received June 17, 1948.]

SUMMARY.

A detailed account is given of a quantum-mechanical treatment of the lower cycloparaffins, cyclobutadiene and ethylene. It is found that strained bonds (in the original sense of Baeyer) are to be described as bent. Strain energies comparable with those derived from thermochemical data are calculated; the properties of strained systems are discussed; and, in particular, a detailed description of the bonds in cyclopropane is given. The stability of cyclobutadiene has been considered with respect to that of the related molecules benzene, cyclooctatetraene and diphenylene; and hybridization in ethylene has been re-examined. The pairing approximation is used throughout, but an Appendix is added, in which the molecular orbital method is applied qualitatively to cyclopropane.

§ 1. INTRODUCTION.

UNTIL recently the electronic theory of molecular structure does not appear to have been generally applied to geometrically strained molecules such as cyclobutane. Penney (1934b) has considered the peculiar stability of the benzene ring with respect to a supposedly planar cyclooctatetraene and the unknown cyclobutadiene. But in this work he

* Communicated by the Authors.

tacitly assumed that even strained bonds are straight (the precise significance attached to the words "straight" and "bent" will become clear later). Recently Duffey (1946) has revived this assumption—without apparently realizing that Penney's arguments lead one to describe cyclobutane and cyclobutene, together with cyclobutadiene, as impossibly unstable. Kilpatrick and Spitzer (1946), using Pauling's (1931) definition of bond strength, have shown that strained bonds are more correctly considered as bent. Since, however, they would have to describe a π -bond (extreme case of a bent bond) as having zero strength, a more fundamental approach is clearly required. This need is emphasized by the recent discovery of the surely highly strained spiro-pentane (Donohue, Humphrey and Schomaker 1945) and of diphenylene (Waser and Schomaker 1943), whose existence implies the marginal stability of cyclobutadiene.

A suitable treatment of these molecules has been found by regarding the hybridization ratios of the carbon atom as unknown parameters whose values are to be obtained by use of the variation principle. A preliminary account of this method has already been published elsewhere (Coulson and Moffitt 1947): in this present paper we propose to examine the properties of several types of strained systems as thoroughly as possible.

§ 2. CONCERNING HYBRIDIZATION.

2.1. Hybridization Ratios as Variational Parameters.

The pairing approximation or HLSP theory of the electronic structure of molecules may be formulated, very conveniently, in terms of the vector model: Dirac (1935) has shown how the perturbation energy of binding on bringing together electrons in orbitals ψ_i may, quite generally, be written in the form

$$E = \text{const.} - \frac{1}{2} \sum_i \sum_{j>i} (1 + 4 \overline{s_i \cdot s_j}) \cdot J_{ij},$$

where

$$J_{ij} = \iint \psi_i^*(1) \psi_j^*(2) \mathcal{H} \psi_j(1) \psi_i(2) d\tau_1 d\tau_2.$$

This formula is valid provided that any lack of orthogonality between the normalized ψ_i , and higher exchange integrals may be neglected. The quantities $\overline{s_i \cdot s_j}$ are the eigenvalues of scalar product operators representing the relative orientations of the spin vectors of the electrons associated with the orbitals ψ_i, ψ_j .

For example, if s_i and s_j are antiparallel, then $\overline{s_i \cdot s_j} = -\frac{3}{4}$, and it is said that these electrons are (fully) coupled, or perfectly paired. If s_i and s_j have their spins fully saturated by being perfectly paired with other spin vectors, then their relative orientation will be random and $\overline{s_i \cdot s_j} = 0$. (This result follows when we note that for the triplet state of parallel spins we have $\overline{s_i \cdot s_j} = +\frac{1}{4}$.)

A rigorous solution, within this approximation, for some given molecule would necessitate a determination of the values of these $\overline{s_i \cdot s_j}$ by a

variational treatment such as that outlined by Pauling (1933)—and for such purposes the Dirac model has only a formal utility. Pauling's method consists essentially in representing the complete wave function as a linear sum of functions each of which refers to a particular system of pairing the electrons (set of canonical structures). In any single structure each electron spin is completely paired with one other spin and completely random relative to all the rest. By varying the relative weights of the different structures the correct $\bar{s}_i \cdot \bar{s}_j$ are obtained. This process is described as resonance. However, it frequently happens that there is only one "sensible" way of pairing the electrons, so that the complete set of canonical structures contains only one effective bond pattern. In such a case the molecule may be said to be in a "pure valence state" (PVS). Whenever this, which is called the approximation of perfect pairing, is valid, the spins of electrons in "bonded" orbitals are fully coupled (setting $\bar{s}_i \cdot \bar{s}_j = -\frac{3}{4}$ for these) and those of the pairs of non-bonded electrons are left completely random (by putting $\bar{s}_i \cdot \bar{s}_j = 0$). The energy then becomes

$$E = \text{const.} - \frac{1}{2} \sum_i \sum_{j>i} J_{ij} + \frac{3}{2} \sum_{\text{bonds}} J_{aa'}, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where the first summation is over all pairs of orbitals, as before, and the second is over the bonded pairs $\psi_a, \psi_{a'}$ only.

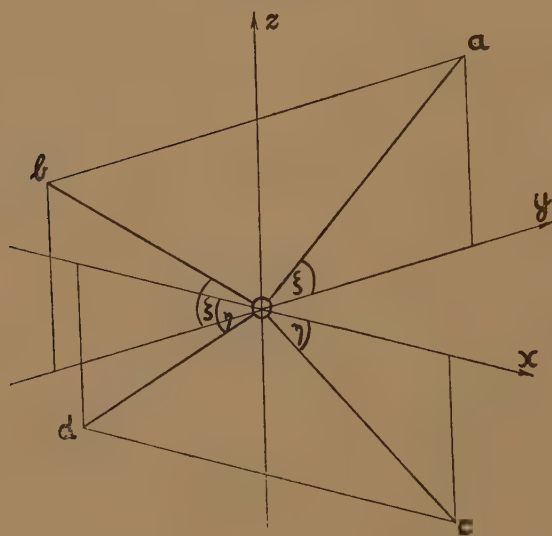
It remains to determine the most appropriate bond pattern for a molecule. This involves the best possible choice of the orbitals $\psi_a, \psi_{a'}$ to be coupled. This may be achieved by replacing the atomic $2s$ and $2p$ orbitals of each atom by orthonormal hybrids. In pictorial language we argue that if a given orbital on atom A' is sometimes paired with a $2s$ orbital of atom A and sometimes with a $2p$ orbital, then we may say that it is "perfectly paired" with some linear combination of $2s$ and $2p$ of A . In this manner hybridization represents a best-possible attempt to represent the molecule by a PVS wave function. But we have still to determine what these linear combinations are. Pauling (1931) determines these by requiring that those orbitals, the spins of whose electrons are to be coupled, shall overlap as much as possible. His criterion has been shown to give the correct "hybridization ratios" for atoms in fields of high symmetry, by means of group theory (Van Vleck 1935, Kimball 1940). For lower symmetries, however, the Pauling procedure is only justifiable if it gives the best possible coupling—and this is sometimes neither obvious nor (*vide infra*) true.

By the variational principle it is clear that the hybridization ratios should really be chosen so as to minimize the energy: only then will the PVS approximation yield useful or apposite results. But whereas it is readily verified that the Pauling condition of maximum overlap gives the same results as the variational treatment for strainless molecules of the open chain type, we shall show that for systems which are geometrically constrained (cyclic molecules) this is not always so.

2.2. Hybridization in Fields of Symmetry C_{2v} .

We have said that in cases of geometrical strain it may not be possible to fulfill the condition of maximum overlapping for every pair of coupled orbitals. This leads to a compromise in which hybridization ratios play the rôle of variational parameters. (The fact that such a compromise is necessary implies the existence of greater resonance, as will be shown later). We can illustrate this most easily by considering the values of the hybridization ratios for the bonding orbitals of a quadrivalent carbon atom in a field of local symmetry C_{2v} . An example would be CH_2D_2 or cyclobutane. This is the so-called w -model of Van Vleck (1933).

Fig. 1.



Hybridization sp^3 in fields of symmetry C_{2v} .

Oz is the two-fold axis,

zOx, zOy are the planes of symmetry.

A carbon atom of this kind will be in a valence state (some combination of sp^3), so that the unperturbed atomic orbitals which we combine in orthonormal hybrids are $\psi(2s)$, $\psi(2p\sigma_x)$, $\psi(2p\sigma_y)$ and $\psi(2p\sigma_z)$. These hybrids will be directed along four axes Oi ($i=a, b, c, d$, as in fig. 1).

Two of the axes, Oa and Ob , make an angle $\frac{\pi}{2} - \xi$ with the two-fold axis of symmetry, here taken to be Oz . The other two axes, Oc and Od , make an angle $\frac{\pi}{2} - \gamma$ with the downward direction of Oz . As the planes Oab , Ocd are at right angles, we may take them to be the yz and zx planes respectively. It is supposed that the four original atomic orbitals are

$rf(r)$, $xf(r)$, $yf(r)$, $zf(r)$. This means that the four hybrid orbitals may be written in the form

$$\begin{aligned}\psi_a &= l\{\psi(2s) + \lambda\psi(2p\sigma_a)\}, \\ \psi_b &= l\{\psi(2s) + \lambda\psi(2p\sigma_b)\}, \\ \psi_c &= m\{\psi(2s) + \mu\psi(2p\sigma_c)\}, \\ \psi_d &= m\{\psi(2s) + \mu\psi(2p\sigma_d)\},\end{aligned}$$

where $\psi(2p\sigma_a)$ is a $2p$ orbital directed along Oa , and is thus a linear sum of the basic $2p$ orbitals: and where l , m , λ , μ are related to ξ and η by the conditions for orthogonality and normalization:

$$\begin{aligned}1 - \mu^2 \cos 2\eta &= 1 - \lambda^2 \cos 2\xi = 1 - \lambda\mu \sin \xi \sin \eta = 0, \\ l^2(1 + \lambda^2) &= 1 = m^2(1 + \mu^2).\end{aligned}$$

It should be noticed that these equations are only consistent for those values of ξ , η in the range

$$0 \leq |\xi|, |\eta| \leq \pi/4;$$

also that only one independent choice from among the six quantities l , m , λ , μ , ξ , η may be made. For example, let a given (permissible) value be assigned to the quantity ξ ; then the others may be expressed in terms of ξ as follows:

$$\begin{aligned}\lambda &= \sqrt{\sec 2\xi}, & l &= \frac{1}{\sqrt{2}} \sec \xi \sqrt{\cos 2\xi}, \\ \mu &= A \cdot \operatorname{cosec} \xi, & m &= \frac{1}{\sqrt{2}} \tan \xi, \\ \eta &= \sec^{-1}(A \cdot \sec \xi), & A &= \sqrt{(\cos^2 \xi + \cos 2\xi)}.\end{aligned}$$

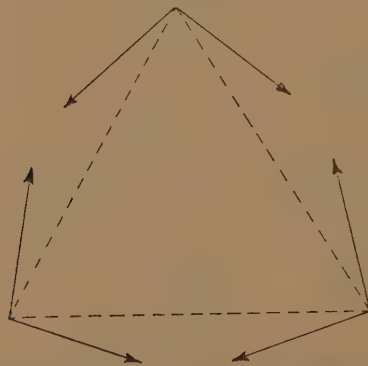
If $\xi = \eta = \sin^{-1} \frac{1}{\sqrt{3}}$, the symmetry becomes tetrahedral T_d for the central carbon atom, and the hybridization ratios λ and μ are each $\sqrt{3}$. This is the well-known tetrahedral system of equivalent orbitals. Alternatively, if $\xi = \pi/4$, so that the bonds Oa and Ob are at right angles, then $\eta = 0$. This implies that these bonds are of pure $2p$ type.

It must be emphasized that these orbitals do not necessarily point at the atoms which they bind; so that although the pairs of equivalent atoms bound lie in the planes of symmetry aOb , cOd respectively, they are not in general on the lines Oi ($i=a, b, c, d$) as was tacitly assumed by Penney. Indeed, they will certainly lie off these lines, if, as in cyclic systems, there is any geometrical constraint on the attached atoms arising from the rest of the molecule. That is to say, bonds are in general "bent" and only straight for strainless molecules (*vide infra*). It is interesting to observe that if we wish to use the language of electron pair bonds, we are obliged to describe cyclopropane by means of bent bonds. This is because the angle $(\pi - 2\xi)$ between two hybrid orbitals Oa , Ob , cannot be decreased to a value less than $\pi/2$, and the angle between the directions of the adjacent carbon atoms is only 60° . As

a result, the hybrid orbitals responsible for binding the carbons must point in directions, shown by the arrows in fig. 2, which are not along the straight lines joining the atoms. In this way the bonds are "bent."

We may now use equation (1) to write down the energy associated with a particular carbon atom, whose environment in a given molecule is specified and of the required symmetry. This energy may be written in terms of only one parameter, which may conveniently be taken to be ξ . This function $E(\xi)$ must now be minimized with respect to ξ in order to obtain the best description of our carbon atom compatible with the approximation given by a PVS model.

Fig. 2.



Arrows denote the directions of the hybrid orbitals in the carbon plane of cyclopropane.

2.3. Pauling's Criterion.

Before proceeding, however, it is useful to examine in greater detail the condition of maximum overlap. In Pauling's (1931) original scheme it is stated that

"Of two eigenfunctions (Hybrid orbitals) with the same radial dependence, the one with the larger value in the bond direction will give rise to a stronger bond, and for a given eigenfunction, the bond will tend to be formed in the direction with the largest value of the eigenfunction."

According to what we have said before, this last postulate can only be applied in the absence of strain. However, in his treatment of the relative stability of the cyclic molecules $(CH)_n$, Penney (1934b) has tacitly assumed that the first clause is equivalent to the condition that we must choose hybrid orbitals with maxima in the directions of the incipient bonds. But, as Kilpatrick and Spitzer (1946) have pointed out, this is by no means the same as saying, with Pauling, that the bond strength is proportional to the value of the hybrid orbital in the bond direction. For although a given hybrid orbital may attain its maximum

value in some other direction than that of the bond it may nevertheless have a greater value in the bond direction than that particular hybrid whose maximum does lie in this direction.

As an example of this, let carbon atoms lie on the lines $z=y$, $z=-y$ of fig. 1 with $z>0$; this would be the case for cyclobutane. Then if the bonding orbitals have maxima (*i. e.* are directed) along these lines, as Duffey (1946) or the arguments of Penney would have us believe, then, because the lines are perpendicular, the orbitals would need to be of pure $2p$ type. Let us use Slater functions for the atomic orbitals, so that

$$\psi(2s) = Nre^{-cr} = F(r), \text{ say,}$$

and

$$\psi(2p) = N\sqrt{3r} \cos \theta e^{-cr},$$

where N is a normalizing factor. Then the orbitals referred to earlier have values $\sqrt{3}F(r) = 1.732 F(r)$ in the bond directions. But Kilpatrick and Spitzer have shown that the hybrid orbitals with $\xi = 37^\circ 15'$ have the greatest numerical values in the bond directions, despite the fact that their maxima lie along directions Oa , Ob which differ by $45^\circ - 37^\circ 15' = 7^\circ 45'$ from the bond directions. These maximum values are numerically equal to $(1 + \lambda^2)^{-\frac{1}{2}}(1 + \sqrt{3}\lambda \cos \pi/4 - \xi)F(r) = 1.984 F(r)$, which is considerably greater than the former $1.732 F(r)$.

The assumption embodied in Pauling's postulate is readily justified when dealing with "straight" $\sigma\sigma$ -type bonds. As the strain increases, however, it becomes less applicable until finally it breaks down completely for unsaturated molecules (regarding ethylene as the limit, $n=2$, of the series of cyclic hydrocarbons $(CH_2)_n$): for the "bond strength" of any $\pi\pi$ bond, in this sense, would be zero, since the bond direction lies in the nodal plane of the bonding π orbitals. We may compare Penney's (1934a) work on ethylene, in which he demonstrates the superiority of the Hund-Mulliken model to that of Pauling's original formulation. In a similar way we shall find that Kilpatrick and Spitzer's results are more acceptable for cyclopentane than for cyclopropane.

Guy (1946) has also discussed the concept of hybridization in HLSP theory and shown that it may be regarded as a particular form of resonance using non-hybridized atomic functions—the orthogonality of the hybrid functions in the PVS yields certain conditions which restrict the free variation of the parameters with respect to which the energy of the molecule is minimized in the full HLSP treatment. However, since he too assumes that the hybridization ratios are determined by the criterion of maximum overlapping, his approach is mainly of theoretical interest.

§ 3. GENERAL TREATMENT OF THE CYCLOPARAFFINS.

3.1. General Formulation of the Problem.

Let us now consider the cycloparaffins $(CH_2)_n$, where $n=2, 3, 4, 5$. Provided that $n \geq 5$ the carbon atoms may be taken to be coplanar and equivalent in each molecule. We need only calculate, therefore, the

The three quantities on the right-hand side are defined as follows: $C_n(\xi)$ is the energy of the bond C-C' and is represented by a "perfect pairing" formula

$$C_n(\xi) = B_c - \frac{1}{2} \sum_{i,j'} J_{ij'} + \frac{3}{2} J_{aa'}.$$

In this formula \sum sums over all the hybrid orbitals ψ_i of atom C and all the hybrid orbitals $\psi_{j'}$ of C', but $J_{aa'}$ is simply the exchange integral between the orbitals ψ_a and $\psi_{a'}$ which are paired together. Since the Coulomb term B_c is independent of spin inclinations, it does not vary with ξ and it is readily shown that the term containing the double summation does not depend on ξ either. $H(\xi)$ represents the C-H₂ energy and is of the same form as $C_n(\xi)$:

$$H(\xi) = B_H - \frac{1}{2} \sum_i J_{Hi} + \frac{3}{2} (J_{Hc} + J_{Hd}).$$

It is assumed that the Coulomb integral here, namely B_H , does not vary greatly over a range of HCH angles.

$A(\xi)$, which contains the atomic terms, has been evaluated by Van Vleck (1934) and will be treated as a constant in the early stages of the calculations.

Now, in terms of the carbon orbitals associated with the axes XYZ, ψ_a may be written

$$\psi_a = l \{ \psi(2s) + \lambda \cos \theta_n \psi(2p_z) + \lambda \sin \theta_n \psi(2p_{\pi_z}) \},$$

where

$$\theta_n(\xi) = \frac{\pi}{n} - \xi, \quad l = \frac{1}{\sqrt{2}} \sec \xi \sqrt{(\cos 2\xi)}, \quad \lambda = \frac{1}{\sqrt{2}} \sec \xi.$$

It is soon found that so far as its variation with ξ is concerned, the energy takes the form

$$E_n(\xi) = \text{const.} + E_n^{\sigma\sigma}(\xi) + E_n^{\sigma\pi}(\xi) + E_n^{\pi\pi}(\xi) + E_n^H(\xi).$$

$E_n^{\sigma\sigma}(\xi)$, $E_n^{\sigma\pi}(\xi)$, $E_n^{\pi\pi}(\xi)$ represent the non-constant contributions of $\sigma\sigma$ -, $\sigma\pi$ - and $\pi\pi$ -type C-C' bonding to the total energy. Similarly $E_n^H(\xi)$ is that part of the C-H₂ bonding energy which is a function of ξ .

The explicit expression for these various contributions are best given in terms of the integrals $C_{\alpha\beta\gamma\delta}$ and $N_{\alpha\beta}$ introduced by Penney (1934a) and defined, since the atomic functions are all real, by

$$C_{\alpha\beta\gamma\delta} = - \iint \psi_c(\alpha_1) \psi_c(\beta_2) \mathcal{H} \psi_c(\gamma_1) \psi_c(\delta_2) d\tau_1 d\tau_2,$$

$$N_{\alpha\beta} = - \iint \psi_H(1s_1) \psi_c(\alpha_2) \mathcal{H} \psi_c(\beta_1) \psi_H(1s_2) d\tau_1 d\tau_2.$$

\mathcal{H} is the usual Hamiltonian operator, and α , β , γ , δ assume the values s , σ , π according as they refer to the atomic $2s$, $2p\sigma$, $2p\pi$ orbitals of carbon.

[A little care is needed to distinguish between σ used as a subscript in $C_{\alpha\beta\gamma\delta}$ or $N_{\alpha\beta}$, and σ used as a superscript as in $E^{\sigma\sigma}$. In the first case-

it refers to an atomic $2p\sigma$ orbital: in the second it refers to a bonding of $\sigma\sigma$ -type, *i. e.* one whose wave function is symmetrical around the bond direction.]

On account of the symmetrical way in which the $C_{\alpha\beta\gamma\delta}$ come into the formulæ it is convenient to employ the abbreviations $C_{(u)}^{\sigma\sigma}$, ($u=0, 1, \dots 4$), $C_{(v)}^{\sigma\pi}$ ($v=0, 1, 2$) and $C^{\pi\pi}$ respectively for Penney's integrals C_{ssss} , $C_{ss\sigma\sigma}$, $\frac{1}{3}(C_{ss\sigma\sigma}+C_{s\sigma s\sigma}+C_{\sigma s s\sigma})$, $C_{ss\sigma\sigma}$, $C_{\sigma\sigma s\sigma}$, $\frac{1}{3}(C_{ss\pi\pi}+C_{s\pi s\pi}+C_{\pi s \pi s})$, $\frac{1}{3}(C_{s\sigma\pi\pi}+C_{\sigma\pi s\pi}+C_{\pi\pi s\sigma})$, $\frac{1}{3}(C_{\sigma\sigma\pi\pi}+C_{\sigma\pi\sigma\pi}+C_{\pi\pi\sigma\sigma})$ and $C_{\pi\pi\pi\pi}$. Explicitly it may be shown that in terms of these functions:—

$$E_n^{\sigma\sigma}(\xi) = -9 \sec^4 \xi \sum_{u=0}^4 \left\{ \frac{1}{u! (4-u)!} \right\} \cos^u \left(\frac{\pi}{n} - \xi \right) \cdot (\cos 2\xi)^{(4-u)/2} \cdot C_{(u)}^{\sigma\sigma},$$

$$E_n^{\sigma\pi}(\xi) = -\frac{3}{2} \sec^4 \xi \sin^2 \left(\frac{\pi}{n} - \xi \right) \sum_{v=0}^2 \left\{ \frac{1}{v! (2-v)!} \right\} \times \cos^v \left(\frac{\pi}{n} - \xi \right) \cdot (\cos 2\xi)^{(2v-2)/2} \cdot C_{(v)}^{\sigma\pi},$$

$$E_n^{\pi\pi}(\xi) = -\frac{3}{8} \sec^4 \xi \sin^4 \left(\frac{\pi}{n} - \xi \right) \cdot C^{\pi\pi},$$

$$E^H(\xi) = -\frac{3}{2} \{ \tan^2 \xi N_{ss} + (1 + \sec^2 \xi \cos 2\xi) N_{\sigma\sigma} + 2 \tan \xi \sqrt{(1 + \sec^2 \xi \cos 2\xi) N_{s\sigma}} \}.$$

If we minimize this $E_n(\xi)$ with respect to ξ , we obtain the best possible description of the molecule on the pure valence state (PVS) model, and may readily calculate the strain energy of any particular ring system (value of n). If $E_n(\xi)$ attains its stationary value when $\xi = \xi_n$, we may also find the corresponding HCH angle ($\pi - 2\eta_n$), since $\eta_n = \eta(\xi_n)$, as in § 2.2.

3.2. Details of the Minimization.

When minimizing the functions $E_n(\xi)$, a graphical method is adopted; $E_n(\xi)$ is plotted against ξ for each value of n and for given values of the interatomic integrals $C_{\alpha\beta\gamma\delta}$, $N_{\alpha\beta}$. Now, although the values of the C-H integrals $N_{\alpha\beta}$ are fairly well known (Penney 1935), those of the $C_{\alpha\beta\gamma\delta}$ are less certain. Bartlett and Furry (1931) have calculated the values of many of them analytically using simplified atomic functions, but Penney (1934) places more reliance on considerations of (geometrical) overlap and of Mulliken's diagram relating to the electronic energy levels of diatomic molecules (1932). The following features emerge quite clearly, however, and ensure the maximum of consistency with a wealth of semi-empirical data:

$$1 \leq C_{(u)}^{\sigma\sigma} < C_{(u+1)}^{\sigma\sigma} \leq 3 \text{ e.v.} \quad (u=0, 1, 2, 3).$$

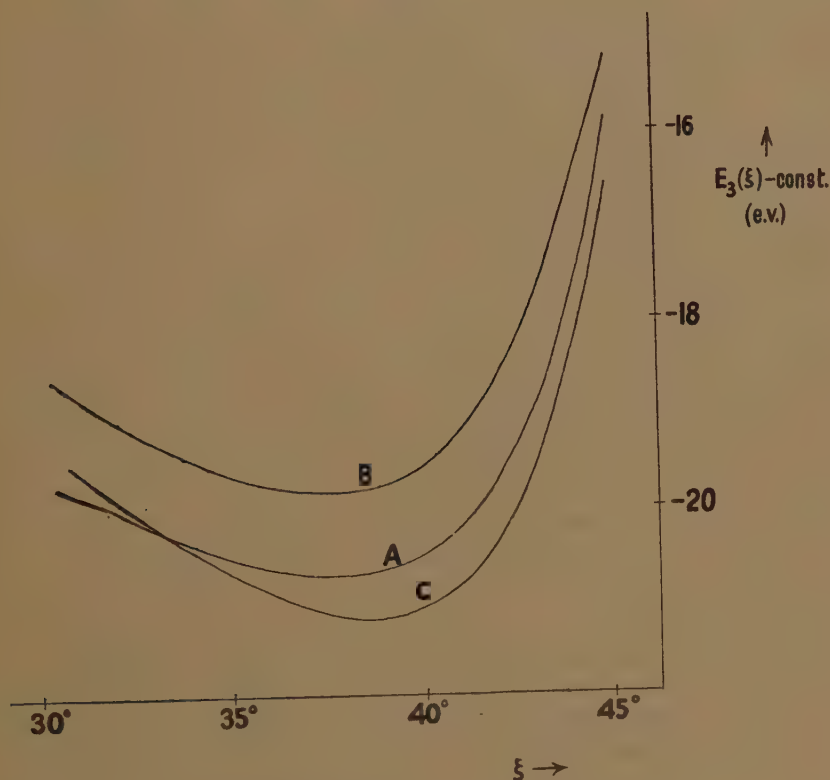
$C_{(2)}^{\sigma\sigma} \simeq 2 \text{ e.v.}$ and $C_{(v)}^{\sigma\pi}$ ($v=0, 1, 2$) small and positive. To these we may add the more reliable values

$$C^{\pi\pi} = 1 \text{ e.v.}; \quad N_{ss} = 2.0, \quad N_{s\sigma} = 2.1, \quad N_{\sigma\sigma} = 2.2 \text{ e.v.},$$

given by Duchesne (1943) and Penney (1935). It turns out numerically

that for $n=3, 4$ and 5 the behaviour of $E_n(\xi)$ is almost entirely determined by the C-H₂ terms and those containing $C_{(u)}^{\sigma\sigma}$; but for $n=2$, $C_{(v)}^{\sigma\pi}$ and C^{nn} are rather more significant. We now choose successively various possible sets of $C_{(u)}^{\sigma\sigma}$ which conform to the relations above, and minimize $E_n(\xi)$ for each set. It is found that the positions of the minima of the $E_n(\xi)$ remain almost unaffected by variations in the choice of these sets, so strong is the angular ξ -dependence.

Fig. 4.

Minimization of $E(\xi)$ for cyclopropane.

For example, consider the minimization of $E_3(\xi)$ —that is, the determination of the best possible hybridization for cyclopropane. First the likely set $\{1.50, 1.75, 2.00, 2.25, 2.50 \text{ e.v.}\}$ of $\{C_{(u)}^{\sigma\sigma} (u=0, 1, \dots, 4)\}$ is chosen and $E_3(\xi)$ is plotted against ξ as in curve A of fig. 4. There is a well-defined minimum near $\xi=38^\circ$. Similar curves are then drawn for alternative feasible sets $\{C_{(u)}^{\sigma\sigma}\}$ of which two, namely those with $\{1.00, 1.5, 2.0, 2.5, 3.0\}$ and $\{2.0, 2.0, 2.0, 2.0, 2.0\}$ are shown as B and C respectively in the figure. In all cases the energy function $E_3(\xi)$ is found to possess a fairly pronounced minimum lying somewhere between $\xi=37\frac{1}{2}^\circ$ and $38\frac{1}{2}^\circ$.

The results of these calculations may be tabulated as follows: (The case of ethylene, in the last row, is considered more fully later, when it is shown that this scheme of pairing is, in fact, not very suitable).

		n	ξ_n	η_n	$\theta_n = (\pi - 2\eta_n)$	$\widehat{\text{HCH}}$	λ	μ
Cyclopropane	..	3	38°	32°	22°	116°	2.03	1.51
Cyclobutane	..	4	36°	34½°	9°	111°	1.80	1.67
Cyclopentane	..	5	35¼°	35¼°	0°	109½°	1.73	1.73
"Ethylene"	..	2	40°	28½°	50°	123°	2.40	1.36

It is interesting to note that a PVS treatment of spiropentane necessarily describes the hybridization at the central carbon atom as tetrahedral ($\xi_c = \eta_c$; $\theta_c \simeq 24^\circ$). By symmetry one would expect the interactions between the two cyclopropylidene rings to be small; spiropentane may therefore be treated as fulfilling, except at the central carbon atom, the functions of cyclopropane twice over.

The Table above shows that cyclopropane is in a rather different category from the larger cycloparaffins. For example, the hybridization parameter λ is distinctly larger, showing that the bonds in the plane of the ring (that is, the C-C bonds) have a larger p -character than in a normal tetrahedral bond, for which $\lambda = 1.73$. As a result the C-H bonds have rather less p , as is shown by the smaller value of μ . This effect is much less marked in cyclobutane and has disappeared in cyclopentane, where the angles and hybridizations are effectively tetrahedral. The departures of the hybridization ratios λ, μ from the equal value 1.73 is shown by an increased HCH angle. This increase is only just detectable in cyclobutane, but it is quite significant in cyclopropane.

3.3. Some other Factors.

There are four factors whose exclusion from the above calculations deserves further consideration before we proceed. Two of these tend to decrease ξ : The atomic term $A(\xi)$ of §3.1. assumes its minimum value for the case of tetrahedral hybridization when it is some 1.4 e.v. lower than when $\xi = 45^\circ$. Similarly a more detailed account of $\sigma\pi$ -type bonding terms which were not considered in the minimization process would exhibit a small tendency to decrease ξ . However, there are two other and probably rather more powerful terms which would tend to increase ξ . In the first place the H-H repulsions which would try to open the HCH angle ($\pi - 2\eta$), have been neglected. In the second place, by restricting ourselves to a PVS model, we have expressly excluded any resonance energy other than that implied in our choice of best possible hybridization. Now, the contribution to the total energy which arises from such resonance, is shown in §5 to be a little larger for greater values of ξ , *i. e.*, greater values of the HCH. However, none of these factors is large, so that our ξ 's, derived in one approximation, probably have values very close to those which they would have in a more refined treatment.

§ 4. FURTHER DISCUSSION OF CYCLOPROPANE.

We have shown in § 3.1 that in PVS theory an energy $E(\xi, \theta)$ may be associated with the CH_2 group in any symmetrical molecule $\frac{\text{R}}{\text{R}} > \text{CH}_2$.

In this expression ξ determines the hybridization ratios (and hence also the HCH angle), and θ measures the angle between the directions of the line C-R and of the hybridized orbital on C which forms part of the C-R bond. The RCR angle is numerically equal to $\pi - 2\xi - 2\theta$. If the hybridized orbitals point along the directions of the bonds, $\theta = 0$, and we may describe the bonds as straight. If all the bonds are equivalent, as in methane, $\pi - 2\xi$ is the tetrahedral angle $109^\circ 28'$. In general, neither of these two conditions applies, θ and ξ being determined so as to minimize $E(\xi, \theta)$. But if R is an alkyl group (or some group whose electronegativity is of the same order as that of an alkyl group) and if steric hindrance between the two R groups does not impose geometrical constraints on the RCR angle, the best value of ξ is very near to the tetrahedral value, and $\theta = 0$. So far as ξ is concerned, this result had already been proved by Penney (1935). Not only are such bonds straight; they are also unstrained.

However, in the cycloparaffin series with which we have been concerned, θ is not a free variable. For simple geometrical considerations compel θ to be equal to $\frac{\pi}{n} - \xi$. Thus the energy of any CH_2 group cannot assume its lowest value. The C-C bonds are strained ($\pi - 2\xi \neq 109^\circ 28'$) and they are also bent ($\theta \neq 0$).

The difference between the best unstrained $E(\xi, \theta)$ and the $E_n(\xi)$ which we have obtained by minimization is therefore a measure of the "strain energy" per CH_2 group for each molecule (value of n). This corresponds closely to the thermochemical quantity of the same name. In either case the strain energy is a difference between two relatively large quantities, so that less weight is to be attached to predictions of this than, for example, to the calculations of the HCH angle. (The "thermochemical" strain energy does not correspond to any strictly thermodynamic property of a compound, its value being derived on the basis of the assumption of bond additivity).

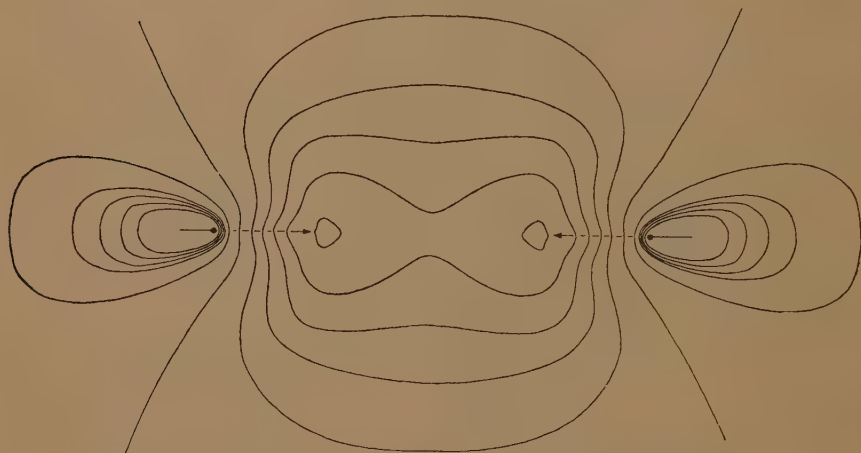
The strain energy per CH_2 group in cyclopropane due both to the different C-H₂ energy and to the changes in $\sigma\sigma$ -type C-C bonding energy, assuming perfect pairing, is some $2\frac{1}{2}$ e.v. Allowing for the (small) contribution of $\sigma\pi$ -type C-C bonding and the additional energy of resonance (shown in paragraph 5 to be approximately 1 e.v.) it is found that the strain energy must lie somewhere between $\frac{1}{2}$ and 1 e.v. per CH_2 group. (Branch and Calvin 1941) give the thermochemical strain energy as $10 \text{ Kals} \simeq \frac{1}{2} \text{ e.v.}$) Owing to the complexity of certain significant refinements which we have not included (*e.g.* an account of the different H-H repulsions would also necessitate a corresponding treatment of the Coulomb integrals, etc.) we shall not try to improve this estimate. From the graphs of fig. 4, it appears that this is an improvement of nearly 5 e.v. per CH_2 ,

group over a model in which the most nearly "straight" ($\xi=45^\circ$) type of hybridization is assumed—Duffey's model—and some 0.2 e.v. better than that ($\xi=39^\circ 45'$) of Kilpatrick and Spitzer.

4.2. Description of the Bonds.

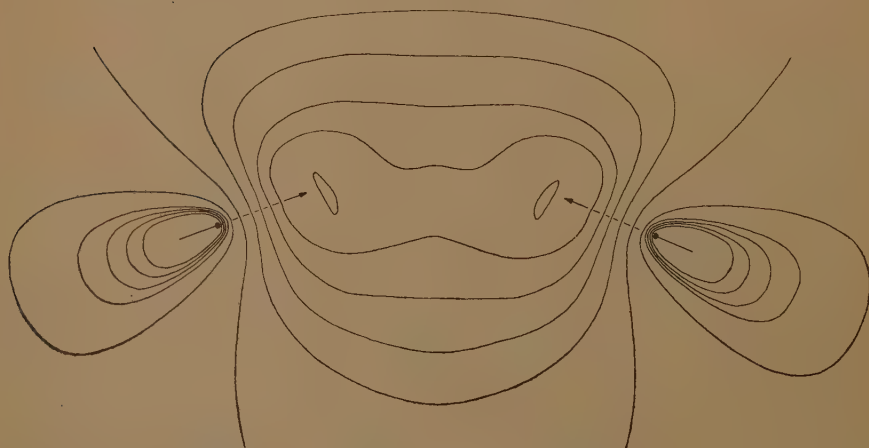
Those bonds which contain hybrid orbitals whose maxima do not lie in the directions of these bonds have been described as strained. In

Fig. 5.



Contours of equal electron density in an unstrained bond.

Fig. 6.



Contours of equal density in a strained bond.

order to amplify this statement the electron density curves for both an unstrained and a strained C-C bond have been drawn and are shown in figs. 5 and 6. A nuclear separation of 2.9 atomic units (1.53 Å) and

Torrance's (1934) self-consistent-field atomic functions for carbon were used; this is not because the calculations were considered to merit the use of more accurate AO's than, say, orthogonalized Slater functions, but because Torrance's functions had already been used to calculate hybrid orbitals on a previous occasion (Moffitt and Coulson 1947). The electron density at any point is given by

$$\frac{\psi_C^2 + 2S\psi_C\psi_{C'} + \psi_{C'}^2}{(1+S^2)},$$

where ψ_C , $\psi_{C'}$ are the bonding orbitals of C, C' respectively, and $S = \int \psi_C \psi_{C'} d\tau$ is the overlap integral.

In the case of the unstrained bond, ψ_C is given explicitly by

$$\psi_C = \frac{1}{2}\{\psi_C(2s) + \sqrt{3}\psi_C(2p\sigma)\},$$

and for the strained bond,

$$\psi_C = l\{\psi_C(2s) + \lambda[\cos \theta \psi_C(2p\sigma) + \sin \theta \psi_C(2p\pi)]\},$$

where

$$l = \frac{1}{\sqrt{2}} \sec \xi \sqrt{(\cos 2\xi)}, \quad \lambda = \frac{1}{\sqrt{2}} \sec \xi, \quad \theta = \frac{\pi}{3} - \xi, \quad \xi = 38^\circ.$$

In both cases similar expressions hold for $\psi_{C'}$.

Contours of equal electronic density are drawn in the plane of the cyclopropane ring for the strained bond. The inter-nuclear axis is not now a symmetry axis. The arrows show the directions of the maxima of the hybrid orbitals; these are the directions at the carbon atom in which the hybrid orbitals may be said to point. For an unstrained bond these coincide with the bond directions, but in the cyclopropane ring they are inclined at an angle $\theta = 60^\circ - 38^\circ = 22^\circ$ to the bond direction. It is well known that throughout a considerable range of variations in "L-S" structure, electron density curves for an atom remain essentially unaltered—so that the refinements of including the results of our fuller HLSP treatment in § 5, while greatly increasing the ardours of computation, would result in curves which are practically indistinguishable from those in figs 5 and 6. The asymmetry in charge distribution for the strained bond is seen to be quite marked.

Because of the interest in the cyclopropane system at the time of writing (*cf.* Walsh 1947, Robinson 1947) the total electron density diagram for all six electrons responsible for C_3 bonding has been given in fig. 7. This set of curves is found by the superposition of the three density diagrams such as that of fig. 6 for each of the bent bonds; it is therefore an aggregate electron distribution which does not show the extent to which the individual bonds contribute. It is noticeable that there is a considerable plateau of charge in the middle of the ring. In passing from one of the carbon atoms towards the centre of the equilateral triangle which they form, the high electron density at the atom is seen to fall off rapidly into a trough, before rising again onto the central plateau.

Also, in the regions where the density of the C-H bonding electrons is high, that of the C₃ bonding electrons is seen to be low—justifying the approximation of neglecting their mutual interactions.

The first and most obvious quantitative prediction to which one is led, is that of the angle between the two straight C-H bonds in cyclopropane. Duffey gave this as 180°, but Kilpatrick and Spitzer preferred 122°. The present calculations give HCH=116° (§ 3.2). A recent and very careful electron diffraction study of the molecule (Bastiansen and

Fig. 7.



Contours of equal electron density for the six C₃ bonding electrons in the plane of the cyclopropane ring.

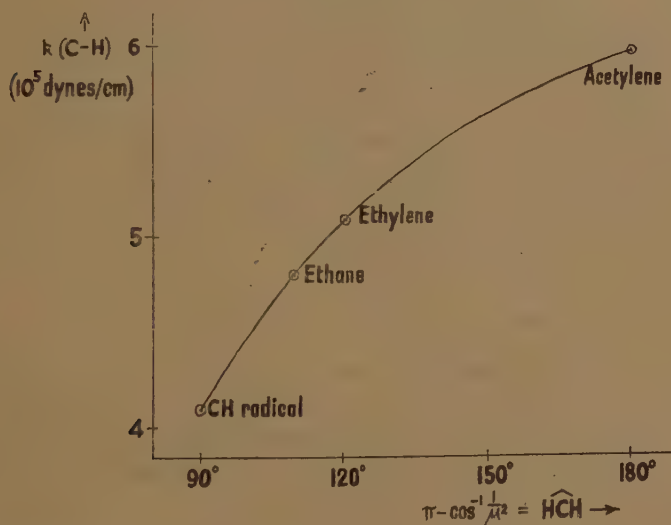
Hassel 1946) has revealed the presence of two sharp maxima corresponding to C-H distances of 1.08 Å and 2.24 Å; since the C-C maximum is at 1.535 Å, this gives an HCH angle of 118° to within $\pm 2^\circ$, lending powerful support to the arguments of paragraph 3.

The C-H bonds in cyclopropane are expected, on the basis of their hybridization ratios μ , to be more closely akin to those of ethylene than those of ethane. But those in cyclobutane are more similar to those of ethane or cyclopentane :

	$\widehat{\text{HCH}}$	CH force constants	μ
Cyclopropane	116°	$5.0 \cdot 10^5$ dynes/cm.	1.51
Cyclobutane	111°	4.85	1.67
Cyclopentane	$109\frac{1}{2}^\circ$	4.8	1.73
Ethane	$109\frac{1}{3}^\circ$	(4.8)	1.73
Ethylene	120°	(5.1)	1.41
*Acetylene	" 180° "	(6.0)	1.00
*CH radical	" 90° "	(4.1)	—

We may use the table above to estimate force constants for these CH bonds. For if we accept the C-H stretching force constants shown in brackets for the last four molecules of the table, we may draw a curve (fig. 8) relating force constant and hybridization ratio μ . This curve enables us to predict the force constants for the three cycloparaffins

Fig. 8.



Relation between force constant of the C-H bond and the HCH angle.

shown at the top of the table. Our value for cyclopropane is in perfect agreement with that (5.0) given by Linnett (1947), and the spectral assignments of Wilson (1943) are consistent with our figure for cyclobutane. (With their HCH angle of 122° for cyclopropane, Kilpatrick and Spitzer would have had to conclude that the C-H force constant in this molecule was greater than that in ethylene).

If the behaviour of C-H bonds with varying hybridization ratio applies also to C-C bonds, we should expect them to have lower force constants in the strained systems than in a normal unstrained tetrahedral C-C bond. This is because $\lambda=2.03$ for cyclopropane and $\lambda=1.73$ for the normal bond. The calculations of Saksena (1939) and Wilson (1943) confirm this conclusion both for this molecule and for cyclobutane, and for ethane (Stitt 1939).

* The HCH angles have only a formal significance for these molecules.

4.3. Bond Lengths.

We have seen that strained bonds are energetically less stable than their unstrained counterparts, and that the force constants which have been associated with the C-C valence stretchings of such bent bonds are less than their analogues in a straight bond such as ethane. Now "weak" bonds are generally longer than strong bonds, so we should at first have expected the C-C bond lengths in cyclopropane and spiropentane to be appreciably greater than that in ethane (1.55 Å).

However, the electron diffraction experiments of Donohue, Humphrey and Schomaker (1945), as well as those of Bastiansen and Hassel (1946), indicate that instead of a lengthening, a bond shortening is observed. The amount of this shortening is not very clear: The former authors give 1.49 ± 0.01 Å as the average C-C distance in spiropentane and the latter, with an improved technique, claim a significant shortening from their figure of 1.535 Å for cyclopropane. But it does not concern us greatly at the moment whether it is of the order of 0.01 Å or of 0.05 Å—we must try to see how a shortening, however small, could occur instead of the lengthening which usually characterizes weakened bonds.

According to §3.1 the PVS energies representing C-C interactions in the unstrained ethane, and in the strained cyclopropane may be written as functions of the internuclear distance r in the forms

$$E_e(r) = P(r) - \sum_u a_{(u)}^{\sigma\sigma} C_{(u)}^{\sigma\sigma}(r),$$

$$E_s(r) = P(r) - \sum_u b_{(u)}^{\sigma\sigma} C_{(u)}^{\sigma\sigma}(r) - \sum_v b_{(v)}^{\sigma\pi} C_{(v)}^{\sigma\pi}(r) - b^{\pi\pi} C^{\pi\pi},$$

respectively; $P(r)$, a monotonically decreasing function of r , represents the repulsion and is the same in both cases. For equilibrium the first derivative of E with respect to r vanishes, so that if r_e and r_s are the unstrained and strained bond lengths respectively and a dot denotes differentiation with respect to r , it is found that

$$\dot{P}(r_e) = \sum_u a_{(u)}^{\sigma\sigma} \dot{C}_{(u)}^{\sigma\sigma}(r_e),$$

$$\dot{P}(r_s) = \sum_u b_{(u)}^{\sigma\sigma} \dot{C}_{(u)}^{\sigma\sigma}(r_s) + \sum_v b_{(v)}^{\sigma\pi} \dot{C}_{(v)}^{\sigma\pi}(r_s) + b^{\pi\pi} \dot{C}^{\pi\pi}(r_s).$$

We are supposing that the hybridization ratios remain unaltered by small changes in r . From this it is clear that

$$\Delta P_{e3} \equiv P(r_e) - P(r_s) = \sum_u (a_{(u)}^{\sigma\sigma} - b_{(u)}^{\sigma\sigma}) \dot{C}_{(u)}^{\sigma\sigma}(\bar{r}) - \sum_v b_{(v)}^{\sigma\pi} \dot{C}_{(v)}^{\sigma\pi}(\bar{r}) - b^{\pi\pi} \dot{C}^{\pi\pi}(\bar{r}),$$

where \bar{r} lies somewhere between r_s and r_e . Assuming, as is reasonable, that $\dot{P}(r)$ is also a monotonically increasing function of r , it is clear that $r_e < \text{or} > r_s$ according as $\Delta P_{e3} < \text{or} > 0$ respectively.

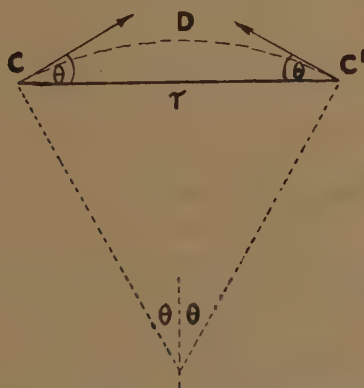
If we were to neglect the contribution of $\sigma\pi$ - and $\pi\pi$ -type bonding, and consider only the $\sigma\sigma$ -bonding, we should find that, since $a_{(u)}^{\sigma\sigma} > b_{(u)}^{\sigma\sigma}$ and $C_{(u)}^{\sigma\sigma}(\bar{r}) < 0$ for all μ , $\Delta P_{e3} < 0$ and so $r_e < r_s$ —i. e. the weaker strained bond would be longer than the stronger unstrained bond. This, indeed,

is the theoretical justification for the usual statement that weak bonds are longer than strong bonds. But there are good reasons for supposing that in our case $-\dot{C}^{\pi\pi}(\bar{r})$ and the $-\dot{C}_{(v)}^{\sigma\pi}(\bar{r})$ are greater than the $-\dot{C}_{(u)}^{\sigma\sigma}(\bar{r})$: in such an event the weak bonds would be shorter instead of longer.

The argument rests upon the fact that even very small admixtures of double ($\pi\pi$ -) bond character have an appreciable shortening effect on basically single ($\sigma\sigma$ -) carbon-carbon bonds. This is revealed experimentally in the considerable shortening of the central C-C bond in butadiene, or in Crawford and Brinkley's (1941) correlation of the properties of various $\sigma\sigma$ -type C-C bonds with Duchesne's (1943) values for the variation of $C^{\pi\pi}$ with r . Alternatively, we may consult the purely theoretical calculations of Bartlett and Furry (1931) which show that those integrals which involve $2p\pi$ atomic orbitals vary more rapidly with internuclear distance than those referring to $2s$ and $2p\sigma$ AO's only. For that reason it becomes very plausible that in the formula for ΔP_{e3} the terms in $b^{\sigma\pi}$ and $b^{\pi\pi}$ will exceed those in $(a^{\sigma\sigma}-b^{\sigma\sigma})$, with the result that

$$\Delta P_{e3} \geq 0.$$

Fig. 9.



Geometrical relation between internuclear distance CC' and "bond length" CDC' .

Even if we cannot in this way predict the actual degree of shortening, we can at least understand why the strained bonds of cyclopropane are slightly shorter than the straight C-C bond of ethane. Further, it may similarly be shown that

$$\Delta P_{ef} > \Delta P_{e3} > \Delta P_{e4} > \Delta P_{e5} = 0,$$

implying that

$$r_f < r_3 < r_4 < r_5 = r_e,$$

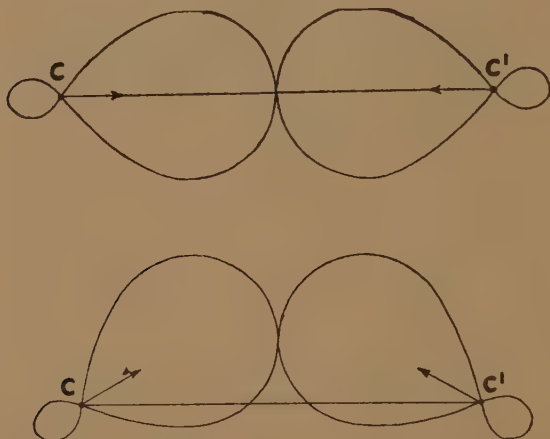
where the suffixes f , 4 and 5 refer to spiropentane, cyclobutane and cyclopentane respectively.

One way of interpreting this result is to distinguish between internuclear distance and bond length; *i. e.* between the length r of the straight line CC' (fig. 9) and the length $r\theta \operatorname{cosec} \theta (>r)$ of the arc CDC' . (N.B.—When

$\theta=22^\circ$, we find $\theta \operatorname{cosec} \theta=1.025$). This means that the length of the strained bond may be longer than 1.55 \AA , although the internuclear distance is less than this. An idea somewhat similar to this has been used by Bernstein (1947) to explain the variation of C-C bond length with bond order.

An alternative pictorial interpretation may be given if we use the conventional representation in which a given hybrid orbital is considered as a hard solid of revolution (similar to the combination of spherical harmonics shown in polar diagrams). Then on account of the strongly directional character of these hybrids, the corresponding solids of revolution are very far from being spherical. It is clear from fig. 10 that the

Fig. 10.



(a) Straight bond.

(b) Bent bond, showing reduced internuclear distance.

two nuclei C, C' can approach each other more closely when the bond is bent. In fact, for any fixed shape of the orbital (value of the hybridization ratio), the internuclear separation is maximum for the unstrained or straight bond between these two nuclei.

§ 5. RESONANCE IN CYCLOPROPANE.

5.1. Calculation of the Resonance Energy.

Van Vleck (1934) has calculated that in methane the error accompanying the PVS approximation—as distinct from the full HLSP treatment—accounts for some $\frac{1}{2}$ e.v. per molecule; *i. e.* about 0.1 e.v. per bond. This is the amount by which resonance lowers the energy below that associated with the best possible scheme of perfect pairing. It may be assumed that, as regards resonance between C-H and either C-C' or other C-H bonding orbitals, his figure of 0.1 e.v. per bond is valid for the cycloparaffins also. But it is not likely to be true for the C-C' bonding orbitals in the plane of a cyclopropane ring. For there are six of these orbitals,

and as fig. 2 shows, none of these orbitals are ever farther from one another than adjacent atoms. Also, on account of the direction of the orbitals, the overlap associated with the best possible PVS discussed in §4 is not so overwhelmingly better than the overlap associated with some other pairing scheme, as it is in unstrained molecules. Consequently one might expect a much fuller resonance as a result of the different ways of pairing these six orbitals, and a correspondingly greater error in the energy from our previous neglect of this resonance.

The full HLSP treatment may be carried through in the usual way by writing down any set of $6!/(3!4!)=5$ independent structures and solving the resulting secular equation. For this purpose we adopt the following notation (see also fig. 13 in the Appendix): The symbol τ_j^i denotes that orbital of C_j which, in the PVS treatment, has been regarded as binding C_i ; $\tau(j:k)$ represents a coupling between (*i.e.* an anti-parallel spin orientation of) the electrons associated with orbitals τ_j^i and τ_l^k in some structure, and (τ_j^i/τ_l^k) is the exchange integral

$$\iint \tau_j^i(1)\tau_l^k(2)\mathcal{H}\tau_l^k(1)\tau_j^i(2) dv_1 dv_2$$

between these two orbitals. We now choose five independent structures as follows: the first four Ψ_P , Ψ_A , Ψ_B , Ψ_C , are defined by the coupling schemes:

$$\begin{aligned}\Psi_P & (\tau_1^6 : \tau_1^2) & (\tau_3^2 : \tau_3^3) & (\tau_1^3 : \tau_3^1), \\ \Psi_A & (\tau_2^1 : \tau_1^2) & (\tau_3^2 : \tau_1^3) & (\tau_2^3 : \tau_3^1), \\ \Psi_B & (\tau_2^1 : \tau_1^3) & (\tau_3^2 : \tau_2^3) & (\tau_1^2 : \tau_3^1), \\ \Psi_C & (\tau_2^1 : \tau_3^2) & (\tau_1^2 : \tau_3^3) & (\tau_1^3 : \tau_3^1).\end{aligned}$$

Thus Ψ_P represents that structure corresponding to the PVS approximation, and the other three are less favourable structures with only one "good bond." The fifth structure may be taken to be

$$\Psi_D \quad (\tau_2^1 : \tau_2^3) \quad (\tau_3^2 : \tau_3^1) \quad (\tau_1^3 : \tau_1^2)$$

as it involves no "good bonds" its contribution to the final wave function may safely be neglected. For the ground state of the molecule, symmetry considerations then enable us to write down the final wave function in the form:

$$\psi = \Psi_P + g(\Psi_A + \Psi_B + \Psi_C),$$

where g is some constant. The secular equation is a simple quadratic. (If, following Pauling (1933), a canonical set chosen with the aid of Rumer's theorem were used, no such simple reduction of the secular problem would be possible).

The elements of the energy matrix are now easily obtained in terms of four primitive exchange integrals, each of which is a function of the hybridization parameter ξ , and a constant quantity containing all those Coulomb and exchange terms which are independent of the state of hybridization of the participating carbon atoms. As before, the position

of the hydrogen atoms determines the value of ξ and conversely. Thus in the usual notation, where \mathcal{H} is the Hamiltonian, E is the energy and R the constant quantity :

$$(P|\mathcal{H}-E|P)=R-E+3\alpha,$$

$$(P|\mathcal{H}-E|A)=\frac{1}{2}(R-E+3\alpha+3\beta-\frac{3}{2}\gamma-\frac{3}{2}\delta), \text{ etc.}$$

$$(A|\mathcal{H}-E|A)=R-E+3\beta, \text{ etc.}$$

$$(A|\mathcal{H}-E|B)=\frac{1}{4}(R-E+3\alpha+3\beta-\frac{3}{2}\gamma-\frac{3}{2}\delta), \text{ etc.}$$

α, β, γ and δ are the four integrals which depend on ξ , and are defined by

$$\alpha=(\tau_j^i/\tau_i^i), \quad \beta=(\tau_j^i/\tau_i^k), \quad \gamma=(\tau_j^j/\tau_i^k), \quad \delta=(\tau_j^k/\tau_i^k), \quad (i \neq j \neq k).$$

Of these, α, β, δ may be expressed linearly in terms of the interatomic C-integrals of § 3, with coefficients which are trigonometric functions of the angle, by writing

$$\tau_j^i=l\psi_j(2s)+l\lambda\{\cos\theta\psi_j(2p\sigma)+\sin\theta\psi_j(2p\pi)\},$$

$$\tau_i^j=l\psi_i(2s)+l\lambda\{\cos\theta\psi_i(2p\sigma)+\sin\theta\psi_i(2p\pi)\},$$

$$\tau_i^k=l\psi_i(2s)+l\lambda\{\cos\phi\psi_i(2p\sigma)-\sin\phi\psi_i(2p\pi)\},$$

$$\tau_j^k=l\psi_j(2s)+l\lambda\{\cos\phi\psi_j(2p\sigma)-\sin\phi\psi_j(2p\pi)\},$$

$$\phi=\theta+\frac{\pi}{3}, \quad \theta\equiv\theta_3=\frac{\pi}{3}-\xi, \quad l=\frac{1}{\sqrt{2}}\sec\xi\sqrt{(\cos 2\xi)}, \quad l\lambda=\frac{1}{\sqrt{2}}\sec\xi.$$

Here σ and π refer to C_i-C_j as axis of quantization. γ is a similar function of ξ and the intra-atomic integrals ; in Van Vleck's notation (1934)

$$\gamma=l^4K(2s;2s)+(\lambda l)^4\{\cos^2 2\xi K(2p\sigma;2p\sigma)+\sin^2 2\xi K(2p\sigma;2p\pi)\} \\ +2l^2(\lambda l)^2\{K(2s;2p)-\cos 2\xi[C(2s;2p)+K(2s;2p)]\}.$$

By inspection it is found that the coefficient $\frac{1}{4}\sec^4\xi\sin^2 2\xi$ of $K(2p\sigma;2p\pi)$ is alone appreciable for our range of ξ , so that this term of the expansion gives γ its monotonically increasing character with these values of ξ . We use Beardsley's (1932) values for the atomic integrals, since these have received strong spectroscopic support (see for example Van Vleck 1934).

Accordingly we find that with ξ near 38°

$$\alpha\simeq-6, \quad \beta\simeq 0, \quad \gamma\simeq+10, \quad \delta\simeq-2 \text{ e.v.,}$$

and the secular equation, with $x=R-E$, becomes

$$\begin{vmatrix} x-18 & \frac{3}{2}(x-30) \\ \frac{3}{2}(x-30) & 3[x+\frac{1}{2}(x-30)] \end{vmatrix}=0,$$

i. e.

$$x=-2\pm 23.3,$$

or, in the ground state,

$$E=R-21.3 \text{ e.v.}$$

The resonance energy E_r is clearly the difference between E and the PVS energy $E_p \equiv (P | \mathcal{H} | P) = R + 3\alpha$:

$$E_r = E_p - E = (R - 18) - (R - 21.3) = 3.3 \text{ e.v.}$$

Thus the resonance energy per CH_2 group is 1.1 e.v.; *i. e.* the additional energy due to the peculiar resonance possibilities in the plane of the cyclopropane ring is $1.1 - 0.1 = 1$ e.v. per CH_2 group. It is to be noticed that, as ξ increases, so does γ and hence also E_r (a result which has been quoted above in § 3.3).

The geometrical factors to which the strain in cyclopropane is attributed are therefore also responsible for the much larger resonance energy of this molecule. That is, the compromise which has to be effected between the competing demands of the various pairs of hybrid orbitals for maximum bonding strength (in PVS) leads to a correspondingly greater degree of delocalization of the participating electrons. The analogous calculation is more difficult to carry through in the higher cycloparaffins but, on account of the considerably reduced bending of the bonds in these molecules, the resonance energy per CH_2 group is much lower. Indeed, in cyclopentane it probably differs negligibly from the value 0.1 as found in unstrained hydrocarbons.

5.2. Manifestations of Resonance.

We can now see how the relatively high degree of delocalization which we have been led to associate with the C-C bonding electrons manifests itself in the behaviour of cyclopropane. We should expect such effects to include a capacity for conjugating with neighbouring unsaturated groups of favourable symmetry as well as for giving the cyclopropyl system an electrophilic character.

As is well known, the latter effect is realized in the low dipole moment of cyclopropyl chloride (1.76 D—Rogers and Roberts 1946), as compared with that of cyclopentyl chloride (2.04 D). (The dipole moment of chlorobenzene, for comparison, is 1.60 D.) Also the spectroscopic shifts to the red of the dinitrophenyl-hydrazones of the cyclopropyl ketones, relatively to those of other saturated ketones (Roberts and Green 1946), and of the derivatives of cyclopropyl ethylene, both give evidence for an unusual conjugating power in a formally saturated group.

§ 6. ETHYLENE, CYCLOBUTADIENE AND CYCLO-OCTATETRÈNE.

6.1. Ethylene.

In dealing with the cycloparaffins $(\text{CH}_2)_n$ with $n \geq 3$, it was found that the particular form of hybridization described in § 2.2 was the only reasonable one for a PVS model. This was also the original scheme proposed for ethylene, implying that the carbon atoms in ethylene were joined by two bent bonds (see fig. 12 (a)). But there is another important alternative to consider: For since the molecule is planar the C atomic orbitals may be chosen so that they belong to one or other of the two

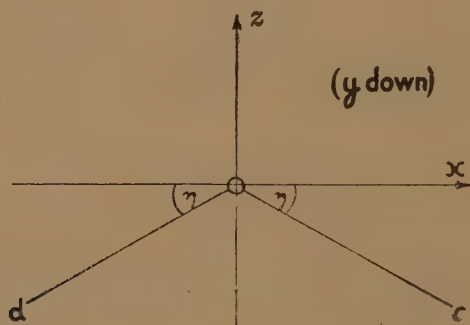
symmetry types, σ - or π -, according as they are symmetrical or anti-symmetrical with respect to reflections in this plane. The following scheme of hybridization is therefore possible, in which the second carbon atom lies along the z -axis of the first (see fig. 11).

$$\begin{aligned} \sigma\text{-type orbitals: } & \begin{cases} \psi_z = n\{\psi(2s) + \nu\psi(2p\sigma_z)\}, \\ \psi_c = m\{\psi(2s) + \mu\psi(2p\sigma_c)\}, \\ \psi_d = m\{\psi(2s) + \mu\psi(2p\sigma_d)\}, \end{cases} \\ \pi\text{-type orbital: } & \psi_y = \psi(2p\sigma_y) = \psi(2p\pi_{xy}). \end{aligned}$$

By normalization and orthogonality

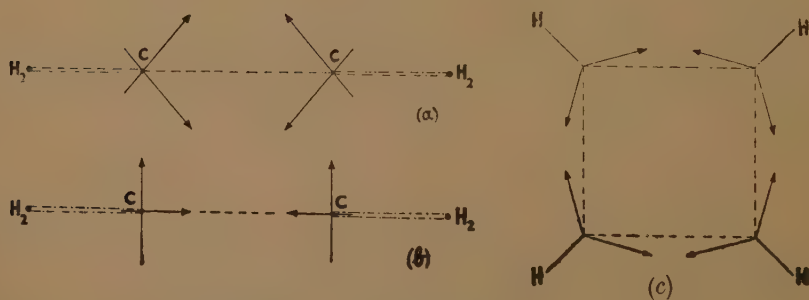
$$\begin{aligned} n^2(1 + \nu^2) &= 1 = m^2(1 + \mu^2), \\ 1 - \mu^2 \cos 2\eta &= 0 = 1 - \nu\mu \sin \eta. \end{aligned}$$

Fig. 11.



Hybridization of carbon atom (at the point O) in ethylene. The axis of y is downwards into the paper.

Fig. 12.



(a) Ethylene, bent bond model.

(b) Ethylene, Hund model.

(c) Cyclobutadiene, bent bond model.

A carbon atom C at the origin would now be described as forming bonds with two hydrogen atoms on the axes Oc, Od and with an equivalent

carbon atom C' lying on Oz , by means of σ -type hybrid bonding orbitals; the binding between the carbon atoms is further increased by coupling the π -orbital ψ_y of C with ψ'_y of C' (see fig. 12 (b)). In this system of pairing ethylene contains four $\sigma\sigma$ C-H bonds, one $\sigma\sigma$ C-C bond and one $\pi\pi$ C-C bond. (Such a hybridization pattern would not lead us to a satisfactory PVS treatment for the cycloparaffins: For since there is no natural way of coupling the orbitals of different atoms, the full resonance to be expected would render a PVS approximation of this kind inappropriate). Penney (1943a) has shown that, for ethylene, the best value of η lies very near $\pi/6$; *i. e.*

$$\mu = \sqrt{2} = \eta \quad \text{and} \quad \sqrt{3}m = 1 = \sqrt{3}n.$$

This may be called the Hund-Mulliken model for ethylene (fig. 12 (b)).

It must now be decided, on a basis of energy considerations, whether it is a better approximation to consider carbon-carbon bonding in ethylene as due to a pair of equivalent bent bonds as in fig. 12 (a), or to one $\sigma\sigma$ and one $\pi\pi$ bond as in fig. 12 (b). Empirical studies all declare in favour of the latter model. Penney has demonstrated its superiority to that originally proposed by Pauling, in which the tetrahedral carbon atom was preserved; but this is not the best (and therefore not a truly representative) bent bond model. As has been shown in § 3, the best bent bond model has an angle $180^\circ - 2 \times 40^\circ = 100^\circ$ between the strained bonds. This is considerably less than $109\frac{1}{2}^\circ$. Comparing the energy of this new model with that of the Hund-Mulliken model on the assumption that the exchange integrals do not vary significantly with changes in the C-C distance, it is found that the latter is more stable by some 6 e.v. By making allowances for the appreciably reduced C-C internuclear distance (see § 4.3 above), the superiority of the π bond model is accentuated—as is to be expected.

Kilpatrick and Spitzer have calculated that the best bent bond model as judged by Pauling's criterion gives a total bond strength of $2\sqrt{2} = 2.83$. The combined strengths of the $\sigma\sigma$ and $\pi\pi$ bonds of Hund's model, on this definition, are $\frac{1}{\sqrt{3}}(1 + \sqrt{2}\sqrt{3})$ and 0, *i. e.* 1.99. This is considerably less than 2.83 although it gives stronger bonding. This exposes the weakness of Pauling's postulate when applied to highly strained or unsaturated systems.

6.2. Cyclobutadiene.

The compound cyclobutene has been known for some time and recently the existence of the dibenz-derivative of cyclobutadiene has been demonstrated. This is difficult to reconcile with Penney's arguments (1934b) which showed that the cyclobutadiene system was enormously strained. As remarked above in § 2, he supposed the σ -electrons of cyclobutadiene to form straight bonds. Let us therefore consider a bent bond model for this hypothetical molecule, taken to have the plane square configuration

shown in fig. 12 (c). The unsaturation or π -electrons have been treated separately elsewhere (for example, Wheland 1938); and these will be ignored at first.

As with the cycloparaffins, the symmetry of the molecule enables one to write down the energy per CH group of the σ -electrons in PVS approximation as a function of some parameter η describing the state of hybridization at each carbon atom. This hybridization may be specified, with reference to fig. 11, as follows: at each carbon atom C, a right-handed set of mutually perpendicular axes Ox , Oy , Oz may be set up so that C is at the origin; Oy is parallel to the principal axis of molecular symmetry and the hydrogen atom H, which C binds, lies along Oz . The carbon atoms C' , C'' adjacent to C are symmetrically disposed with respect to Oz in the xz -plane (z negative). The scheme of pairing is now as in fig. 12 (c). The angle between the directions of the $C \rightarrow C'$, $C \rightarrow C''$ bonding orbitals and the axis of x is taken as the parameter η . This angle, which Penney had assumed to be $\pi/4$, is less than $\pi/4$ when the C-C bonds are bent.

The energy may now be written as a function of η in the following manner:

$$E(\eta) = \text{const.} + C_4(\eta) + H'(\eta);$$

the function C_4 is the same as that defined in § 3.1, but H' differs from H :

$$H'(\eta) = -\frac{3}{2}(\tan^2 \eta \cdot N_{ss} + \sec^2 \eta \cos 2\eta \cdot N_{\sigma\sigma} + 2 \sec \eta \tan \eta \sqrt{(\cos 2\eta) N_{\sigma\sigma}}).$$

When $E(\eta)$ is minimized in the usual way, it is found that the best scheme of hybridization corresponds to $\eta \simeq 33^\circ$. This means that there is an angle of $45^\circ - 33^\circ = 12^\circ$ between the directions of the C-C bonding hybrids and the corresponding bond directions. If we compare the minimum value of E thus obtained with the σ -electron energy in the unstrained benzene ($\eta = 30^\circ$) we find that the strain energy per CH group in this PVS state approximation is about 0.8 e.v. This strain energy associated with the σ -electrons of cyclobutadiene, though greater than that per CH_2 group in cyclobutane, is of the same order and rather less than that in cyclopropane. It is worth noting that although our bonds are not so bent as in cyclopropane or spiropentane, the model we have chosen is some 40 e.v. more stable than that discussed by Penney.

The existence of cyclobutene and diphenylene as stable compounds is now no longer surprising. Indeed it still remains to account for the non-existence of cyclobutadiene, which Penney had set out to explain: If the molecule really is unstable, then the source of its instability cannot be attributed solely to its σ -electrons, for these are very similar to those of stable cyclobutane in the C_4 ring (corresponding to $\theta_4 = 9^\circ$, as found for the latter in § 3.2, we have 12° for the former). We must therefore re-examine the π -electrons; and for this purpose we shall find it much more simple to use the LCAO MO method than the pairing approximation (see Wheland 1938).

The reducible representation whose basic vectors are the atomic functions $\psi_i(2p\sigma_y)$, ($i=1, \dots, 4$), has irreducible components

$$\Gamma = A_u + B_u + E_g,$$

where the notation is that appropriate to a molecule belonging to the point group D_{4h} . Thus there are two non-degenerate and one degenerate molecular orbitals available. Of these A_u is strongly bonding, B_u strongly antibonding and E_g non-bonding; in the ground state, the electronic configuration of the molecule is therefore degenerate and may be written in the form

$$\dots [a_u]^2 [e_g]^2.$$

And, as Wheland (1938) and Coulson (1939) have shown, this implies that the hypothetical molecule would have a triplet ground state (as is obvious in the next higher approximation). There is an evident analogy here between cyclobutadiene and the oxygen molecule, whose ground state has a degenerate outer shell of electrons $[\pi + \pi]^4 [\pi - \pi]^2$; both are paramagnetic.

But Jahn and Teller have proved that stability and orbital degeneracy are simultaneously incompatible unless a molecule is linear (1937, 1938). This means that if the molecule were square, there would be at least one normal mode of vibration with negative restoring force constant, showing that this nuclear configuration is unstable. The oxygen molecule, however, being linear, is stable. It is now easy to see why cyclobutadiene should be unstable: For since the square arrangement of lowest strain will possess an intrinsic decomposition coordinate, any attempts at its synthesis will lead to the formation of its decomposition products, stable unsaturated hydrocarbons.

In diphenylene, the fourfold symmetry of cyclobutadiene is destroyed and thus the degeneracy is removed. This molecule, whose π -electrons contribute an appreciable resonance energy to the bonding energy, is therefore quite stable.

6.3. Cyclo-octatetraene.

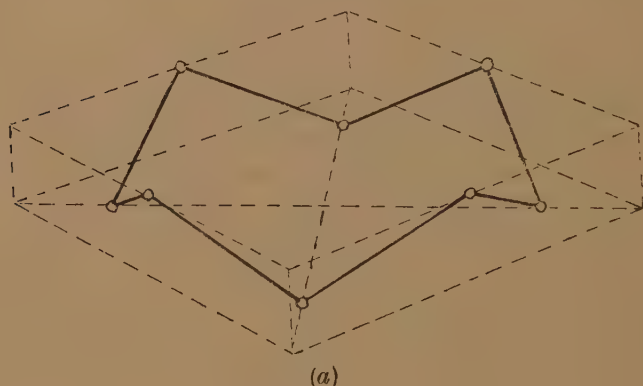
Similar arguments may be applied to cyclo-octatetraene. Unlike cyclobutadiene, however, this molecule has two feasible alternative (non-planar) configurations involving no strain: (fig. 13). The first of these, whose C-C bonds are all equivalent, has four-fold symmetry, whereas the second has two-fold symmetry only and its C-C bonds are alternatively short and long (*vide infra*).

We shall assume the electrons in the "straight" C-C and C-H bonds to be perfectly paired and the hybridization at each carbon atom to be trigonal with $\widehat{CCC} = {}^2_3\pi = \widehat{CCH}$. By virtue of the local symmetry thus gained, we may continue to speak of $\sigma\sigma$ C-C bonds and a π -electron system, though it must be remembered that the quantization of the individual π -orbitals is generally referred to non-parallel axes. The directions of these π -orbitals are chosen so that on flattening the ring

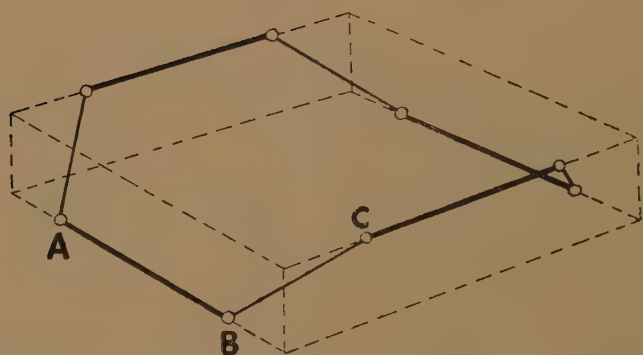
systems they are all parallel to the axis of molecular symmetry. The total $\sigma\sigma$ -bond energy of both models will be the same if variations of the exchange integrals with bond length are neglected. Accordingly, we may decide between our two models by considering the energy contributed by the π -electrons alone.

Now it is readily shown by solid geometry that in the puckered ring model the angle ($\simeq 80^\circ$) between the directions of adjacent π -orbitals

Fig. 13.



(a)



(b)

Cyclo-octatetraene (a) Puckered ring model D_{4d} .(b) Cradle model D_{2d} .

has cosine $(3-2\sqrt{2})$. Since the interaction between adjacent π -orbitals varies as the square of this quantity, it is clear that the π -electron system here is essentially non-bonding. In the cradle type structure, however, adjacent π -orbitals are alternatively either parallel (*e.g.* C_A and C_B) or inclined at an angle ($\simeq 71^\circ$) (*e.g.* C_B and C_C) whose cosine is $\frac{1}{3}$. So although the interaction between C_A and C_B is measured by $C^{\pi\pi}$, that between C_B and C_C is small. Obviously the best PVS scheme is attained by pairing the four pairs of adjacent π_A, π_B orbitals. Thus whereas the π -electrons of the cradle-type D_{2d} model contribute some $4(1-\frac{1}{2} \cdot \frac{1}{9}) |C^{\pi\pi}|$ to the bond energy, those of the puckered ring type add but little to the

stability of the D_{4d} model: The superiority of the cradle type structure is therefore of the order 5 e.v. per molecule. This analysis also holds good for hybridization angles which do not differ greatly from $2\pi/3$.

The PVS approximation therefore leads us to assign cyclo-octatetraene to the point group D_{2d} . It consists essentially of four ethylene units connected by single bonds in a cradle-type structure. The small resonance interaction between these units leads us to expect the molecule to be considerably more reactive than benzene. Our conclusions are in agreement with a substantial amount of experimental evidence. Flett, Cave, Vago and Thompson's spectroscopic studies (1947) seem to indicate that the molecule belongs to the point group D_{2d} . Further, the full resonance which a structure such as D_{4d} would imply, has been disproved by the diamagnetic susceptibility measurements of Pink and Ubbelohde (1947), who have, however, shown that a structure such as our cradle type D_{2d} accounts very well for the observed value. An apparent disagreement exists with the experiments of Bastiansen, Hassel and Langseth (1947), who claim that their electron diffraction examination of the molecule yields only one C-C bond length, rather larger than that in benzene. They have, however, given no details of their analysis. In our model we should have expected the bond lengths to alternate, the shorter double bonds having lengths near 1.34 Å and the longer, single, bonds having lengths lying somewhere near 1.54 Å. This is in perfect agreement with the X-ray diffraction pattern recently observed for this molecule by Kaufman, Fankuchen and Mark (1948). It is of interest that Karle and Brockway have found that tetraphenylene (the tetrabenz-derivative of cyclo-octatetraene) is derived from such a cradle-type structure, with the centres of the four benzene rings at the corners of a tetrahedron; in the same paper (1944), they cite a result of Fieser's which indicates that the dibenz-derivative behaves towards bromine as though it possessed two unconjugated ethylenic double bonds. Finally, it has been shown thermochemically by Prosen, Johnson and Rossini (1947) that cyclo-octatetraene is some 34 Kals./mole less stable than styrene—showing that the π -electronic structure of the former may be well described by means of four strong and practically independent and localized π -bonds. In view of all this it seems fairly certain that our theoretical analysis of this problem is correct.

APPENDIX

Molecular Orbital Description of Cyclopropane.

In giving qualitative descriptions of the behaviour of electrons in molecules, authors tend to intermingle freely the terminologies of molecular orbital and pairing approximations. As this sometimes leads to misunderstanding, a molecular orbital (m.o.) treatment of cyclopropane is offered to complement the preceding calculations which are based on the pairing approximation. This treatment is, however, strictly qualitative, since no quantitative predictions are made.

We suppose, to begin with, that the electrons in the C-H bonds are in perfectly paired and localized σ -type hybrid orbitals. As we showed in §3, these are of a form intermediate between sp^2 and sp^3 . In these localized m-o, the contribution from the carbon atom is similar to that discussed in §2.2 and may be written :

$$(1+\mu^2)^{-\frac{1}{2}}\{\psi(2s)+\mu\psi(2p\sigma_a)\}.$$

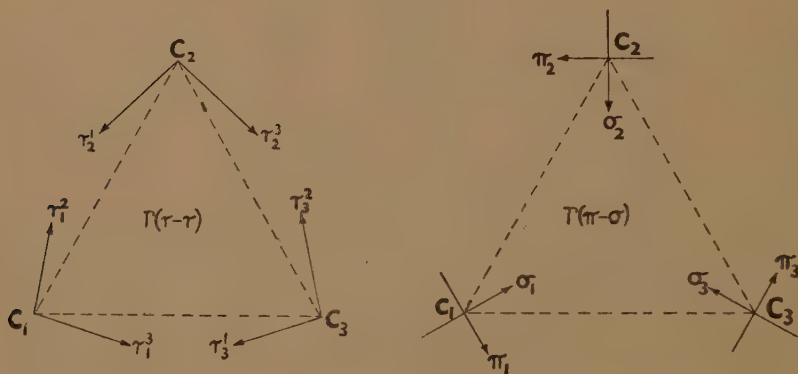
This is equivalent to :

$$(1+\mu^2)^{-\frac{1}{2}}\{\psi(2s)\pm\mu\cos\eta\psi(2p\sigma_x)-\mu\sin\eta\psi(2p\sigma_z)\},$$

where $\eta=\frac{1}{2}\cos^{-1}(1/\mu^2)$ and $\pi-2\eta$ is the HCH angle.

This choice of hybrids for the C-H bonds leaves for our further discussion a total of six orbitals, two from each carbon atom C_i ($i=1, 2, 3$). Any choice of these additional hybrids should conform to the local symmetry C_{2v} around each separate carbon atom. It appears that there are two choices, shown diagrammatically in fig. 14, and which we shall call

Fig. 14.



The two alternative hybridization schemes in cyclopropane.

the $\Gamma(\tau-\tau)$ and $\Gamma(\pi-\sigma)$ representations. In each case the directions of the hybrids, which are shown by the arrows, lie in the plane of the ring. But in $\Gamma(\tau-\tau)$, which is similar to the PVS scheme discussed in §2.2, these hybrids point in equivalent symmetrical directions relative to the equilateral triangle formed by $C_1C_2C_3$. In the notation there introduced, we may write these τ_i^j and τ_i^k where, for example,

$$\begin{aligned}\tau_i^j &= (1+\lambda^2)^{-\frac{1}{2}}\{\psi(2s)+\lambda\psi(2p\sigma_a)\}, \\ &= (1+\lambda^2)^{-\frac{1}{2}}\{\psi(2s)+\lambda\sin\xi\psi(2p\sigma_x)+\lambda\cos\xi\psi(2p\sigma_y)\}.\end{aligned}$$

Here ξ is the angle between the direction of the hybrids τ_i^j, τ_i^k at carbon C_i , and the orthogonality condition gives $\lambda^2=\sec 2\xi$. Finally λ, ξ and η are determined in terms of the one independent parameter μ by the remaining orthogonality condition

$$\lambda\mu\sin\xi\sin\eta=1.$$

In the second representation, which we called $\Gamma(\pi-\sigma)$, the orthogonal hybrids in the plane of the ring are written π_i and σ_i , where, as shown in fig. 14, π_i is a $2p\sigma$ orbital at the vertex C_i directed parallel to the base C_jC_k : and

$$\sigma_i = (1 + \nu^2)^{-1/2} \{ \psi(2s) + \nu \psi(2p\sigma_i) \}.$$

In this last expression $\psi(2p\sigma_i)$ is the $2p\sigma$ orbital at C_i directed towards the centre of the triangle, and ν is a coefficient of mixing which is related to μ, η for the C-H hybrids by the orthogonality relation

$$\mu\nu \sin \eta = 1.$$

Following the usual LCAO MO procedure, linear combinations of those hybrid orbitals of each carbon atom which lie in the plane of the ring are now chosen. Let us consider $\Gamma(\pi-\sigma)$ first. The reducible representation $\Gamma(\pi-\sigma)$ is found to split up according to the scheme

$$\Gamma(\pi-\sigma) = A'_1 + A'_2 + 2E' \quad . \quad . \quad . \quad . \quad . \quad (1)$$

for a molecule belonging to the point group $D_{3h} = D_3 \times \sigma_h$. The corresponding LCAO MO forms are easily shown to be

$$\left. \begin{aligned} [a'_1] &= \sigma_1 + \sigma_2 + \sigma_3, \\ [a'_2] &= \pi_1 + \pi_2 + \pi_3, \\ [e'] &= \begin{cases} (\sigma_1 + t\pi_1) + \omega(\sigma_2 + t\pi_2) + \omega^2(\sigma_3 + t\pi_3), \\ (\sigma_1 + t\pi_1) + \omega^2(\sigma_2 + t\pi_2) + \omega(\sigma_3 + t\pi_3), \end{cases} \\ [\bar{e}'] &= \begin{cases} (t\sigma_1 - \pi_1) + \omega(t\sigma_2 - \pi_2) + \omega^2(t\sigma_3 - \pi_3), \\ (t\sigma_1 - \pi_1) + \omega^2(t\sigma_2 - \pi_2) + \omega(t\sigma_3 - \pi_3), \end{cases} \end{aligned} \right\} . \quad . \quad . \quad (2)$$

where $\omega = e^{2\pi i/3}$ is a cube root of unity and t is a constant found from the secular determinant. Of these orbitals, $[a'_2]$ and $[\bar{e}']$ are strongly antibonding, $[a'_1]$ strongly bonding, and $[e']$ less strongly bonding. In the ground state, therefore, the C_3 bonding electrons of cyclopropane may be described as

$$[a'_1]^2 [e']^4, \quad {}^1A'_1. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

On the other hand, if we use the alternative representation $\Gamma(\tau-\tau)$, we find

$$\Gamma(\tau-\tau) = A'_1 + A'_2 + 2E', \quad . \quad . \quad . \quad . \quad . \quad (1')$$

as before, so that the molecular orbitals are

$$\left. \begin{aligned} [a'_1] &= (\tau_1^2 + \tau_1^3) + (\tau_2^3 + \tau_2^1) + (\tau_3^1 + \tau_3^2), \\ [a'_2] &= (\tau_1^2 - \tau_1^3) + (\tau_2^3 - \tau_2^1) + (\tau_3^1 - \tau_3^2), \\ [e'] &= \begin{cases} (\tau_1^2 + s\tau_1^3) + \omega(\tau_2^3 + s\tau_2^1) + \omega^2(\tau_3^1 + s\tau_3^2), \\ (\tau_1^2 + s\tau_1^3) + \omega^2(\tau_2^3 + s\tau_2^1) + \omega(\tau_3^1 + s\tau_3^2), \end{cases} \\ [\bar{e}'] &= \begin{cases} (s\tau_1^2 - \tau_1^3) + \omega(s\tau_2^3 - \tau_2^1) + \omega^2(s\tau_3^1 - \tau_3^2), \\ (s\tau_1^2 - \tau_1^3) + \omega^2(s\tau_2^3 - \tau_2^1) + \omega(s\tau_3^1 - \tau_3^2), \end{cases} \end{aligned} \right\} . \quad . \quad . \quad (2')$$

s is another constant found from the secular determinant appropriate to this representation. Now it is easy to see that the following relations hold between the two pairs of basic functions :

$$(\tau_i^j + \tau_i^k) \sin \xi = \sigma_i; \quad (\tau_i^j - \tau_i^k) \cos \xi = \pi_i.$$

This shows that the two sets of $[a'_i]$ orbitals in (2) and (2') are identical : and that the sets $[e']$ are also identical if $t(1+s) = (1-s) \tan \xi$. But the constants t, s are to be chosen so as to give the best possible LCAO forms for the degenerate molecular orbitals—and may therefore be regarded as parameters with respect to which the energy is to be minimized. Thus (2), (2') are completely identical. This conclusion is obviously true *ab initio* on more fundamental grounds ; but it appears (*e. g.* Walsh 1947) that this explicit demonstration of the complete equivalence of the two representations $\Gamma(\pi-\sigma)$ and $\Gamma(\tau-\tau)$ may not be redundant.

By suitably varying μ (or η), we may run through the whole range of C-H type hybrids, from ($\mu=1$, $\widehat{\text{HCH}}=180^\circ$ —digonal) through ($\mu=\sqrt{2}$, $\widehat{\text{HCH}}=120^\circ$ —trigonal) and $\mu=\sqrt{3}$, $\widehat{\text{HCH}}=109\frac{1}{2}^\circ$ —tetrahedral) to ($\mu=\infty$, $\widehat{\text{HCH}}=90^\circ$ —pure p). Thus to this degree of approximation, the LCAO MO description of cyclopropane is equivalent to varying the energy of the molecule with respect to the two parameters μ and t (or η and s , for example). The number of integrals whose values are uncertain, however, is such that this mode of treatment is at present numerically impracticable. The best HCH angle must therefore be found by other means, such as our work in § 4.

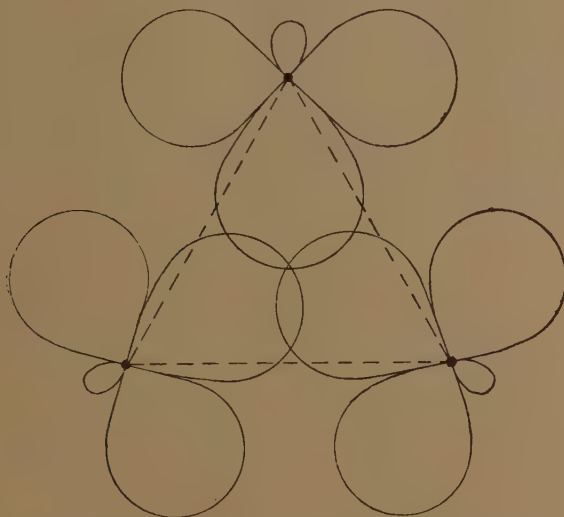
Let us now consider Walsh's (1947) recently published description of cyclopropane in the light of the foregoing analysis. His diagram, which is reproduced in fig. 15, simply means that he prefers us to use the $\Gamma(\pi-\sigma)$ representation—a perfectly legitimate choice even though it has just been shown to be arbitrary. But it is more difficult to reconcile his qualitative description of the molecular orbitals by means of fig. 16 with the above treatment. Two electrons, according to Walsh, are to be assigned to the region a , giving a shell $[a]^2$, say. Two more electrons are supposed to occupy an orbital which gives high probability densities in the zones b, d and e , $[b]^2$, say ; and the remaining electrons are assigned to $[\bar{a}]^2$, a "first overtone" of $[a]$. This implies that the electrons may be separated without degeneracy as follows :

$$[a]^2 [b]^2 [\bar{a}]^2.$$

But although Walsh's $[a]^2$ electrons are clearly our $[a'_i]^2$, his non-degenerate electrons which have been called $[b]^2 [\bar{a}]^2$ do not correspond to the filled (degenerate) shell $[e']^4$ of (3). The two remaining pairs in Walsh's description cannot therefore be considered as moving, in pairs, in independent orbitals, for they do not conform to any allowable symmetry types or irreducible representations. His diagram of fig. 16 should be compared with our calculated total electron density curves of fig. 7.

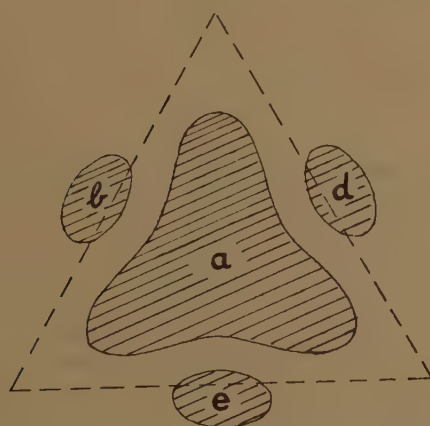
Whether the carbon atoms are to be described as more nearly trigonal or tetrahedral also cannot be determined by a comparison between Walsh's "endwise" and "sideways" overlap of the $\psi(2p\sigma_y)$'s, (*i.e.* the π_i orbitals of the representation $\Gamma(\pi-\sigma)$), but depends essentially upon a

Fig. 15.



$\Gamma(\pi-\sigma)$ diagram for cyclopropane (after Walsh).

Fig. 16.



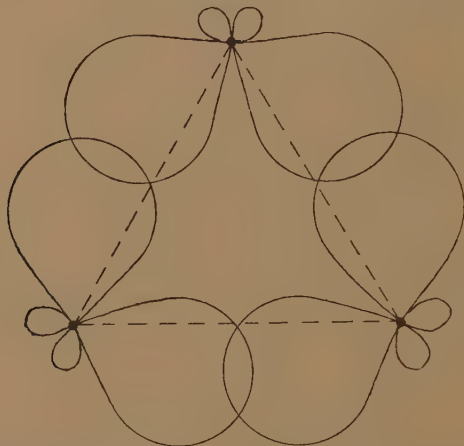
Electron density pattern for cyclopropane (after Walsh).

compromise between the opposing demands for strong C-H and strong C-C bonds. On the contrary, since the $\Gamma(\pi-\sigma)$ and $\Gamma(\tau-\tau)$ representations for both the full HLSP and the LCAC MO treatments are equivalent, overlap considerations only suggest that by analogy with our §5 a PVS model (with perfect pairing) is a good approximation using the

latter (*i. e.* $I(\tau-\tau)$, see fig. 17); but that the representation $I(\pi-\sigma)$ would admit of no such simplification and require a relatively elaborate treatment in which resonance played a much more significant rôle.

The terms "trigonal" and "tetrahedral" in a rigorous treatment of the six C_3 bonding electrons, by either pairing or orbital method, are now only apposite when describing the C-H bonding or the amount of $\psi(2s)$ and $\psi(2p\sigma_z)$ involved in C-C and C-H bonding: Their geometrical significance is restricted to a description of the HCH angle. When, therefore, it is affirmed that the state of the carbon atoms in cyclopropane is more nearly trigonal than tetrahedral, we must interpret this, if it is to mean anything at all, by saying that the HCH angle is to lie nearer 120° than $109\frac{1}{2}^\circ$ and that, correspondingly, μ lies nearer to $\sqrt{2}$ than $\sqrt{3}$.

Fig. 17.



Perfect pairing model for cyclopropane (bent bonds).

The empirically determined HCH angle and the force constant of the C-H bonds, which lie nearer the corresponding figures for ethylene than those for, say, ethane, support these conclusions.

As has been seen, the PVS treatment predicts the HCH angle to be 116° , in entire agreement with Walsh's expectations. We believe, however, that the description of the molecule's C-C bonds as "bent" is much more simple and richer intuitively than is Walsh's more complicated symbolism; and much less liable to lead to misunderstandings. Our method has the additional advantage of lending itself to quantitative treatment.

The conclusion that there is a considerable degree of delocalization in the cyclopropane bond system, accounts for the chemical and spectroscopic conjugating effects rather more clearly than does Walsh's qualitative argument. For the energy of delocalization in cyclopropane is some 1 e.v. per C-C bond, to be compared with a value 5 or 10 times smaller in ethane.

REFERENCES.

- BARTLETT, 1931, *Phys. Rev.*, **37**, 507.
BASTIANSEN, and HASSEL, 1946, *O. Tidsskr. Kemi. Bergv.*, **6**, 71.
BASTIANSEN, HASSEL, and LANGSETH, 1947, *Nature, Lond.*, **160**, 128.
BEARDSLEY, 1932, *Phys. Rev.*, **39**, 913.
BERNSTEIN, 1947, *J. Chem. Phys.*, **15**, 284.
BRANCH, and CALVIN, 1941, *Theory of Organic Chemistry*.
COULSON, 1939, *Proc. Roy. Soc. A*, **169**, 413.
COULSON, and MOFFITT, 1947, *J. Chem. Phys.*, **15**, 151.
CRAWFORD, and BRINLEY, 1941, *Ibid.*, **9**, 69.
DIRAC, 1935, *Principles of Quantum Mechanics*.
DONOHUE, HUMPHREY, and SCHOMAKER, 1945, *J. Am. Chem. Soc.*, **67**, 332.
DUCHESNE, 1943, *Physica*, **10**, 817.
DUFFEY, 1946, *J. Chem. Phys.*, **14**, 342.
FLETT, CAVE, VAGO, and THOMPSON, 1947, *Nature, Lond.*, **159**, 739.
FURRY, and BARTLETT, 1931, *Phys. Rev.*, **38**, 1615; 1932, *Ibid.*, **39**, 210.
GUY, 1946, *Comp. Rend.*, **223**, 670.
JAHN, and TELLER, 1937, *Proc. Roy. Soc. A*, **161**, 220.
JAHN, 1938, *Ibid.*, **A**, **164**, 117.
KARLE, and BROCKWAY, 1944, *J. Am. Chem. Soc.*, **66**, 1974.
KAUFMAN, FANKUCHER, and MARK, 1948, *Nature, Lond.*, **161**, 165.
KILPATRICK, and SPITZER, 1946, *J. Chem. Phys.*, **14**, 463.
KIMBALL, 1940, *J. Chem. Phys.*, **8**, 188.
LINNETT, 1947, *Nature, Lond.*, **160**, 162.
MOFFITT, and COULSON, 1947, *Phil. Mag.* **38**, 634.
MULLIKEN, 1932, *Rev. Mod. Phys.*, **4**, 1.
PAULING, 1931, *J. Am. Chem. Soc.*, **53**, 1367; 1933, *J. Chem. Phys.*, **1**, 280.
PENNEY, 1934a, *Proc. Roy. Soc. A*, **144**, 166; 1934b, *Ibid.*, **A**, **146**, 223; 1935, *Trans. Faraday Soc.*, **31**, 734.
PINK, and UBBELOHDE, 1947, *Nature, Lond.*, **160**, 502.
PROSEN, JOHNSON, and ROSSINI, 1947, *J. Am. Chem. Soc.*, **69**, 2068.
ROBERTS, and GREEN, 1946, *J. Am. Chem. Soc.*, **68**, 214.
ROBINSON, 1947, *Nature, Lond.*, **159**, 400.
ROGERS, and ROBERTS, 1946, *J. Am. Chem. Soc.*, **68**, 843.
SAKSENA, 1939, *Proc. Ind. Acad. Sci.*, **A**, **10**, 449.
STITT, 1939, *J. Chem. Phys.*, **7**, 297.
TORRANCE, 1934, *Phys. Rev.*, **46**, 388.
VAN VLECK, 1933, *J. Chem. Phys.*, **1**, 177, 219; 1934, *Ibid.*, **2**, 20, 297; 1935, *Ibid.*, **3**, 803.
WALSH, 1947, *Nature, Lond.*, **159**, 167, 712.
WASER, and SCHOMAKER, 1943, *J. Am. Chem. Soc.*, **65**, 1451.
WHELAND, 1938, *Proc. Roy. Soc. A*, **164**, 397.
WILSON, 1943, *J. Chem. Phys.*, **11**, 369.

II. *Investigation of Soft Radiations by Proportional Counters* *.—I.

By S. C. CURRAN, J. ANGUS, and A. L. COCKROFT,
Department of Natural Philosophy, University of Glasgow †.

[Received October 12, 1948.]

SUMMARY.

The possibilities of detecting and measuring the energy of soft radiation, particularly β -radiation, γ -radiation and K-capture X-rays, in the range 30 ev. to 150 kev. are discussed and verified experimentally. It is shown that high sensitivity, good resolution and accuracy of energy determination can be obtained through most of this range by combining a suitably designed proportional counter and high gain amplifier of low noise level. The energy required per ion pair formed by electrons of energy between ~ 200 ev. and 50 kev. is found to be nearly constant for argon and nitrogen gases. The K-capture radiation of Cu^{64} and the γ - and X-radiation emitted by Ra D are analysed.

INTRODUCTION.

VARIOUS forms of proportional counting tubes have been studied and used since their initial application by Rutherford and Geiger (1908). They have been used principally to differentiate between the ionization produced by heavily and lightly ionizing radiations, and the full understanding of their mode of action was delayed for many years after their introduction. The explanation of the fundamental mechanism of gas amplification is now fairly clear and various authors have recently (Korff 1946, Corson and Wilson 1948) given useful summaries of the present position.

In the present papers it will be shown that proportional tubes used in conjunction with a strictly linear amplifier of high gain can be applied to the study of many types of radiation with considerable advantages over other techniques. The counters are constructed to incorporate the accepted features of good counter design; the method of using them in conjunction with an amplifier of very low noise level has not, however, been previously explored systematically, and it will be shown that this combination offers a direct and accurate method of studying soft radiations of electromagnetic or corpuscular nature, a study which is difficult by other techniques and which may be impossible when only weak intensities are available.

* The subject matter of this paper was outlined by Prof. P. I. Dee at the Birmingham Conference on Physics (September, 1948).

† Communicated by the Authors.

We shall begin by describing the apparatus, considering

I. the construction of the proportional counters as it affects performance,

II. the design features of the linear amplifier,

III. the method of recording and analysing the pulse output, and shall follow this by indicating the scope of the method of investigation. Thereafter, results of investigations of the gamma- and X-radiation of Radium D will be discussed. In the following paper an analysis of the beta spectrum of H^3 is presented. It will be shown that beta-radiation, gamma-radiation and X-radiation (including that resulting from K-capture) of energy from ~ 100 ev. to ~ 150 kev. can be rapidly and accurately analysed.

GENERAL PRINCIPLES.

The gas multiplication of a proportional counter can be defined as the ratio of the charge on the electrons collected at the wire to the charge on the electrons produced by an ionizing particle passing through the gas. A proportional counting tube, filled with a suitable mixture of gases or vapours, for example, argon and methane, can be operated at a voltage which gives a charge multiplication up to 10^4 ; strict proportionality with almost any mixture of gases and vapours can be maintained up to a multiplication of at least 100 (assuming that electron attachment is negligible). An amplifier of good design can be made to give a pulse output of several times the peak noise level when a charge of 1500 electrons is collected at the first grid. Thus fairly accurate measurements of charge can be made by collecting ~ 5000 electrons at the grid and therefore, if a proportional counter operating at a gas multiplication of 100 is used, the formation of 50 electrons, as ion pairs in the tube, gives an output pulse whose amplitude is readily measurable to an accuracy of a few per cent. Taking the energy loss per ion pair as 30 ev., the production of 50 ion pairs requires an expenditure of energy of 1.5 kev. Clearly, energies much lower than this value may be measured, provided that strict proportionality of gas multiplication is maintained at the higher gas gains necessary. While the limit of energy measurable corresponds to the production of one ion pair, the statistical fluctuation in the numbers of electrons released by very soft radiations become pronounced and consideration must also be made of the statistical variation in the value of the gas amplification. However, if for these regions resolving power is not a critical consideration, good estimates of the energy of ionizing radiation remain feasible.

It is important to note that the collection of 5000 electrons at the sensitive electrode of an ionization chamber attached to a linear amplifier, where no gas multiplication is employed, gives a pulse from which the energy expended by the particle in producing the ion pairs may be determined with reasonable accuracy. Since this involves an energy of the

order of 150 kev., the proportional counter may be regarded as extending the range of ionization measurements down to the lowest energies involved in ion pair production. The complete range from one ion pair upwards is therefore covered by means of

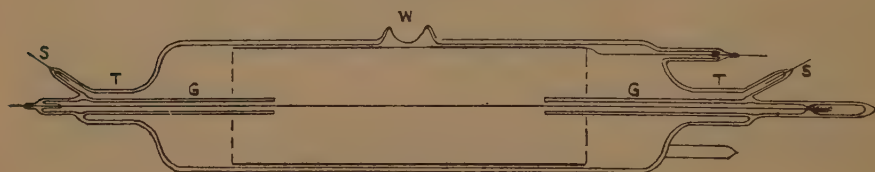
- (a) the proportional counter (~ 30 ev. to 150 kev.),
and (b) the ionization chamber (150 kev. upwards) where each is used in conjunction with an amplifier of high gain and low noise level.

DESCRIPTION OF APPARATUS.

I. *The Construction of Proportional Tubes and Performance.*

The important features of the design are best illustrated by reference to fig. 1, which shows one of the counters used by the authors. The tube is constructed of pyrex glass and is provided with a window, W, about $1/10$ mm. thick to allow easy entry of soft radiations near the middle of the active volume. The cathode (which is maintained at negative

Fig. 1.



Constructional Features of Glass Counter.

high voltage) consists of a thin sheet of smooth aluminium formed into a cylinder and slipped into the tube, the contact to this electrode being brought out through a seal in the envelope. The central wire of tungsten (2 to 4 mil in diameter) is held taut by a spring at one end: the cathode considerably exceeds the exposed part of the wire in length. The central wire is protected at each end by narrow tubes of glass G, which are coated with aquadag on the outside; these aquadag coatings, together with those on the inside and outside of tubes T, are grounded by seals S and serve as guardrings preventing any current from leaking from the cathode to the collecting wire. The aquadag serves a second important purpose; if possible the central wire should be arranged so as not to "see" charged insulating materials, and in these counters this effect is reduced to a low value. The central wire is also protected by a grounded shield over the short length required to connect it to the first grid of the amplifier. To reduce noise to a minimum it is advisable to enclose the counter in a metal box, but some methods of construction make this unessential. Since the wire will pick up any variations in the cathode potential, the high voltage supply must be stabilized and smoothed.

We shall now discuss in more detail certain features which affect the performance of these tubes.

(a) *Sensitivity* : in the detection of gamma- and X-radiation sensitivity is of major importance. In the region below 150 kev. the photoelectric absorption coefficient is relatively large and high efficiency of detection may be achieved. An accurate quantitative analysis of the intensity and energy of the radiation was made possible since the counters used here were designed to give a maximum of absorption in the gas and a minimum absorption at the wall. The latter is achieved by using an aluminium cathode and the former by choosing gases and vapours which have high absorption coefficients in the range of quantum energies involved. Argon may be used for energies between 3.5 kev. and 50 kev. ; for example, at a quantum energy of 10 kev., the mass absorption coefficient of argon is 62 and for a path length of 7 cm. in argon at atmospheric pressure, 54 per cent of the quanta will be absorbed and detected. This represents a very high sensitivity to quanta.

For higher quantum energies, say, around 100 kev., good sensitivity may be obtained by use of gases of high atomic number and by an increase of the counter dimensions ; for example, at this energy, a path length of 10 cm. in xenon at atmospheric pressure gives an efficiency of 15 per cent. At these energies the conditions for obtaining an accurate determination of the quantum energy also include an increase in the size of the counters. Clearly, the linear dimensions of the counter must be large compared with the range of the photoelectrons released in the gas. Also, since the photoelectrons released in the wall material will emerge with greatly reduced energies, it is desirable that such emission should be as small as possible in comparison with the gas absorption. The heavy gases have larger stopping powers and this, by limiting the range, reduces to some extent the dimensions required. As a rough working rule the radius of the counter should not be less than the range of the photoelectrons being investigated.

(b) *Uniformity of Multiplication.*

The multiplication depends principally on the strength of the field in the neighbourhood of the central wire and it is therefore important that the wire should be

(1) of uniform radius, within close tolerances, along its whole counting length : it must be smooth and free from prominences,

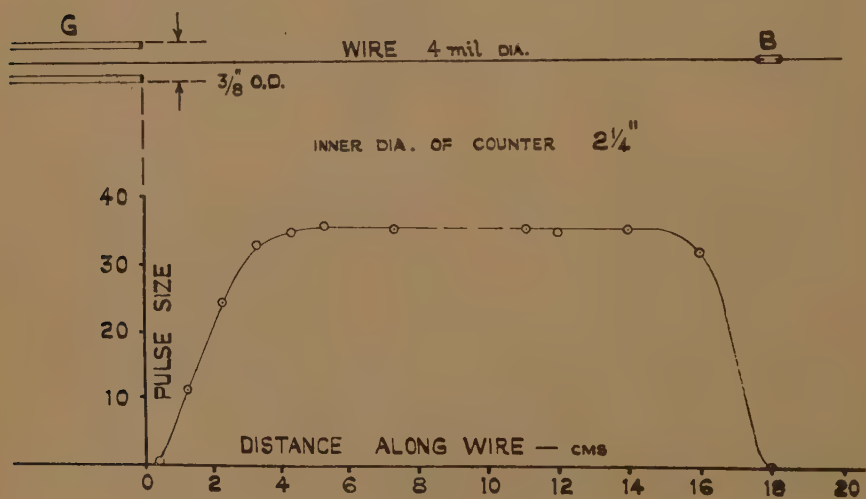
(2) carefully aligned along the axis of the tube, and

(3) arranged to give a minimum end-effect.

With regard to (1), microscopic examination is useful unless the manufacturing tolerances are known. Concerning (2), careless alignment of the wire can lead to erroneous results, since the field strengths on either side of the wire will not be equal. In regard to (3), the field of electrical force varies in intensity and direction near the ends of the wire due to the proximity of the earthed screening tubes.

Some investigations of the latter phenomenon have been published for Geiger Muller counters (Greisen and Nereson 1942) where this distortion gives rise to a variation of efficiency with applied voltage. The form which this phenomenon takes in proportional counters was studied separately here and the results are given in fig. 2. Fluorescence X-rays of copper were passed through a series of thin windows along the length of the counter and histograms of pulse size were obtained for a fixed setting of the tube voltage. The average pulse size was calculated for each window and was regarded as a measure of the gas amplification at various points along the wire. It will be seen from this figure that the relative gas amplification varies rapidly close to the end of the wire, the distance involved in this case being of the order of the radius of the counter. This effect may be neglected for counters designed to detect

Fig. 2.



Variation of gas gain near shield G and bead.

X-rays, since the majority of the photoelectrons will be produced near the window at the centre of the counter; for work with radioactive gases or vapours, however, it is desirable to reduce the end-effect to a minimum. This can be accomplished in various ways, and here we have investigated two methods of effecting a reduction :—

(1) the counter can be made very long compared with the diameter of the cathode, and the shields, G, over the wire (fig. 1) can be made of small radius, say 10 times the wire radius.

(2) the wire may be divided into two unequal portions and these held together by a small glass bead (as indicated at B in fig. 2) to serve as two counters. The two counters are arranged to have unequal lengths, because, since they have similar end-effects, the difference in the histograms of the output pulses obtained when a radioactive element is introduced in gaseous form into the tube will represent that obtained in

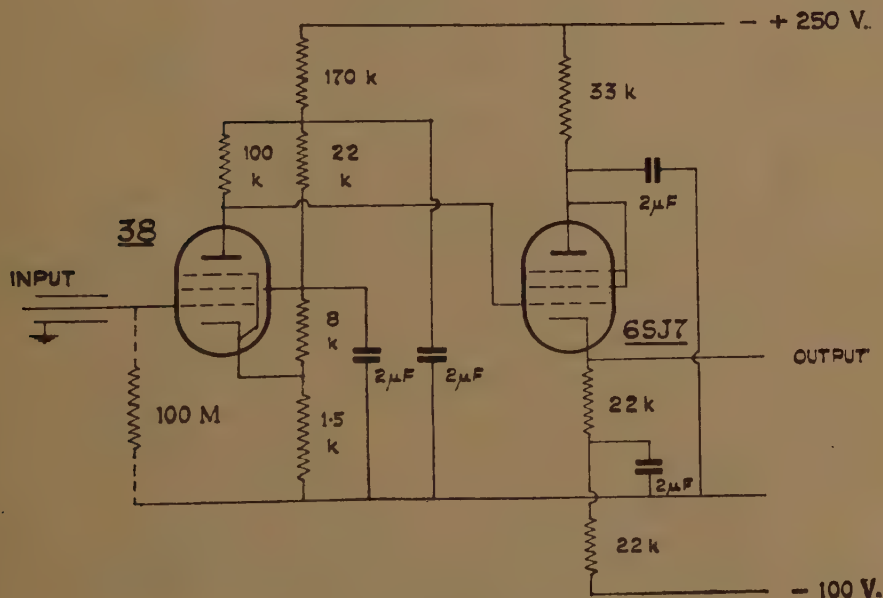
a detector free from end-effect. We show on the right-hand side of fig. 2 the mode of variation of the multiplication in the vicinity of such a bead.

II. The amplifier used with these tubes may be of conventional design, but it is important to realize that very great advantages accrue from the attainment of the best possible signal-to-noise ratio. There has been some tendency to neglect this feature and to expect the gas multiplication to compensate for any inadequacy.

The head amplifier is shown in fig. 3. For maximum sensitivity no grid leak was used, but this occasionally led to difficulty at high counting rates due to biasing-off in the first valve. This difficulty is removed when a 100-megohm resistance is placed in the grid circuit, although some reduction in the signal-to-noise ratio results.

The main amplifier was of the push-pull type throughout.

Fig. 3.



Head amplifier circuit.

III. Throughout this work an oscillographic method of analysing the pulse output was used. The cathode ray tube spot was deflected vertically by the output voltage of the amplifier and the film (35 mm.) was moved horizontally through the camera at a uniform speed; the camera lens had an aperture of $f/1.8$. Counting rates of about 10,000 per minute could be recorded at a film speed of 2 in. per second. This method was chosen in favour of the pulse discriminating circuit method, since it offers a continuous range of determination of the amplitude of the pulses, and because a record of all the pulses necessary for analysis may be obtained in one run of relatively short duration. The requirements of voltage stability of the power unit are considerably relaxed by this technique.

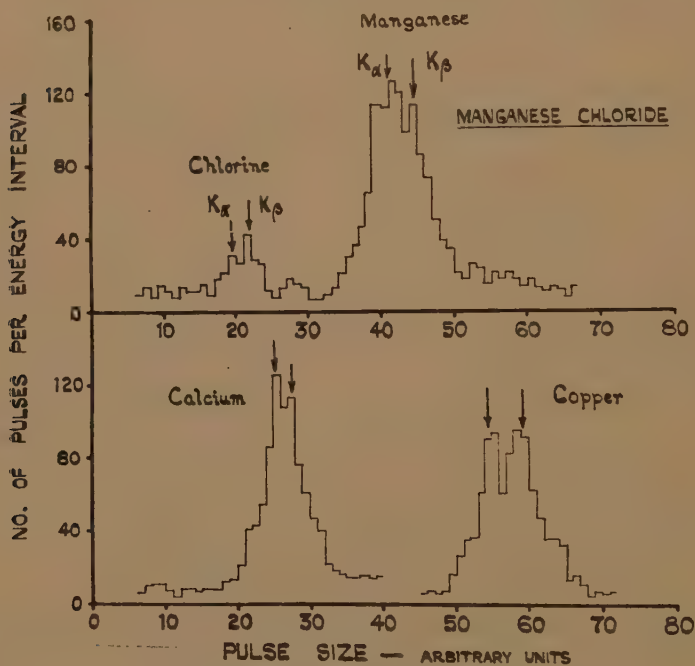
The film is analysed by projecting it upon a screen which is subdivided into 110 amplitude intervals each $\frac{1}{2}$ cm. in height ; pulses falling into each interval are recorded and the histogram of the radiation is built up. This brings us to the question of the calibration of the counter and the scale used in the analysis of the output pulses.

GENERAL PROPERTIES OF THE COUNTERS.

A. *Dependence of the Output Pulse Size on the Energy of the Radiation.*

The variation of output pulse size with energy was studied with X-rays of known energies between 2.6 kev. and 40 kev. An ordinary A.C.-operated X-ray machine of small output intensity emitting white X-rays

Fig. 4.



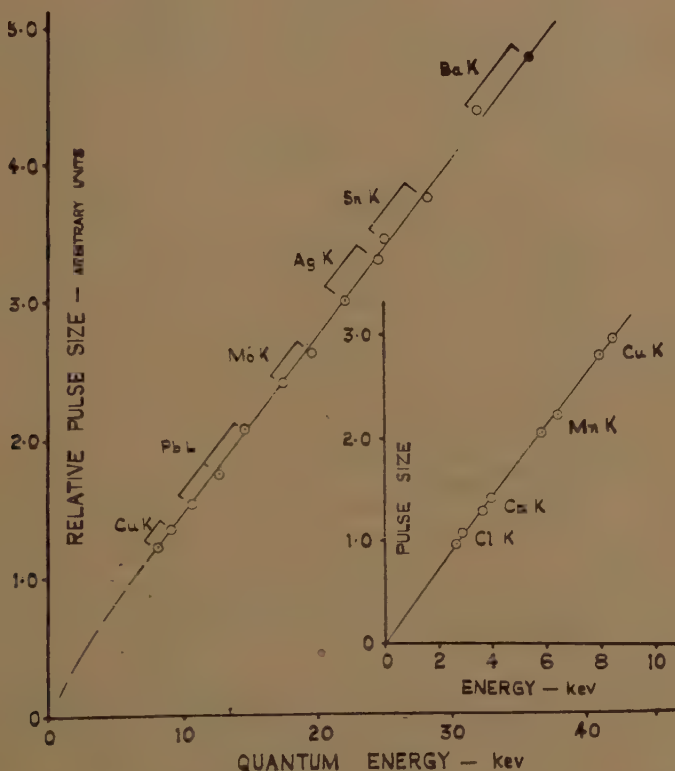
Histograms of fluorescence K X-rays.

was used. The beam fell on a thin scattering foil and the scattered radiation entered the counter at right angles to the beam direction. This scattered radiation was found to consist mainly of the fluorescence X-rays of the scattering element and the histograms obtained showed that it was possible to resolve the K_{α} and K_{β} radiations, or the L_{α} , L_{β} and L_{γ} rays.

Some typical histograms are shown in fig. 4. The average output pulse sizes for these and other characteristic radiations are plotted in fig. 5 against their known quantum energies ; both figures refer to a mixture of argon (60 cm. of Hg) and methane (15 cm. of Hg). It is seen that they lie on or near to a straight line which passes close to the origin.

This indicates that the number of ion pairs formed in the gas mixture is very closely proportional to the energy of the radiation. Since, for these energies, the majority of the quanta will initially produce photoelectrons in the K-level of argon, these photoelectrons will be emitted with energies less than those of the quanta by 3 kev. The X-radiation which would be emitted subsequently by the atoms is very largely internally converted in the outer levels, producing one or more Auger electrons. The Auger effect in argon is about 93 per cent (Compton and Allison 1936) and this high value, together with the fact that the

Fig. 5.



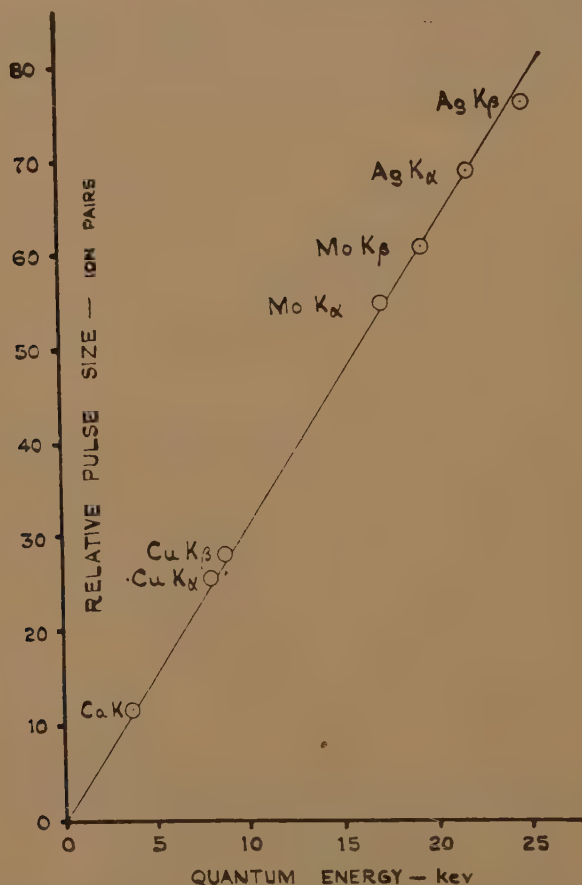
Pulse size as function of quantum energy (for argon and methane).

few X-rays which appear will be rapidly absorbed in neighbouring atoms of the gas, means that in the absorption process the total energy of the photoelectrons and Auger electrons will be very close to the energy of the quantum.

A point of considerable fundamental importance arises from consideration of the relation involved in fig. 5. We have seen that the extrapolated curve passes very close to the origin in accordance with the high Auger effect in argon; the curve also provides strong evidence in favour of the view that the energy expenditure per ion pair by an electron (here a photoelectron) is for argon very nearly constant between 500 ev.

and 50 kev. From a review by Gray (1944) of the results of various investigations of the energy expended per ion pair in air it appears that in this gas this quantity increases at electron energies below 10 kev. Indirect evidence put forward by Gray suggests, however, that the energy expenditure in the noble gases is probably constant down to the limit suggested here. It is a fortunate circumstance in such a low energy region.

Fig. 6.



Variation of pulse size with quantum energy for nitrogen.

We have likewise investigated the variation of pulse size with quantum energy (and therefore with electron energy) for a mixture of nitrogen (60 cm. Hg) and methane (7 cm. Hg) and for methane alone. In the case of nitrogen the variation is remarkably close to linear down to very low energies: this suggests that the change in the energy expenditure per ion pair detected for air (Eisl 1929) is almost entirely due to oxygen.

For methane alone an increase in energy expenditure with increasing energy was found, and the admixture of large proportions of methane should therefore be avoided.

The method used here is well adapted to the study of the variation for many gases and vapours and it is proposed to examine these later. It is difficult to predict the mode of variation of N , the number of ion pairs, with E , the energy of the electron. Thus

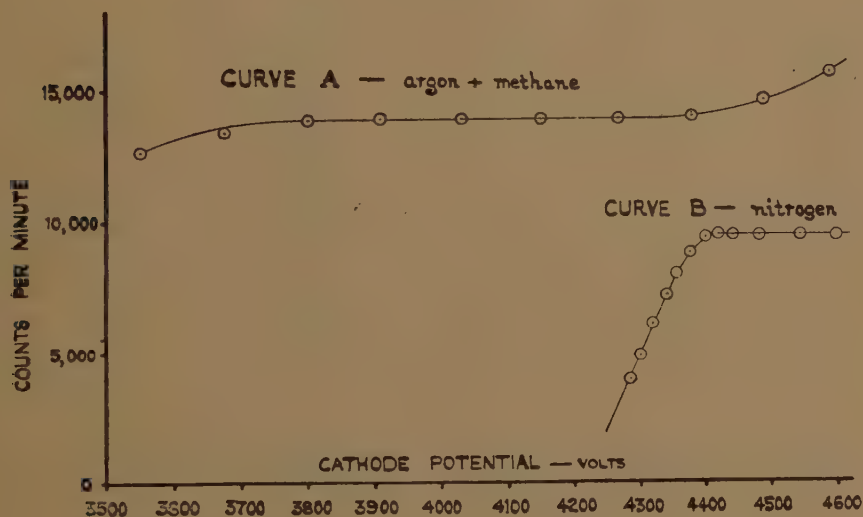
$$dN/dE = (dN/dx)/(dE/dx),$$

and, while dN/dx has been examined experimentally (Smyth 1931) as a function of E for many gases, dE/dx has to be evaluated theoretically. Uncertainties in both quantities make possible large variations of dN/dE as calculated.

B. Quantitative Detection of Radiations.

We have shown that if an electron can form a single ion pair within the counter a multiplication in the gas of the order of 10^3 is adequate

Fig. 7.



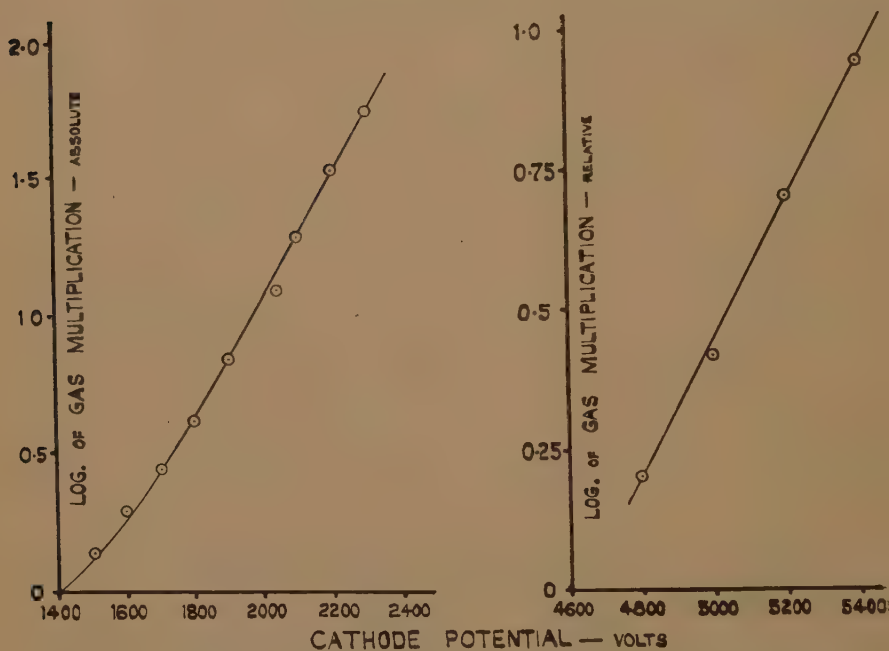
Plateaus in proportional region for β -rays of H^3 (curve A) and X-rays of Cu (curve B).

to give rise to an output pulse from our amplifier sufficiently greater than the noise peaks to permit its recording by a scaler suitably biased to eliminate the noise. A long copper counter, 15 inches useful length and $2\frac{1}{2}$ inches diameter, was filled with a mixture of argon and methane (60 cm. Hg and 15 cm. Hg) and tritium gas, diluted with hydrogen to a partial pressure of about 5 mm. Hg, was added. The rate of counting as a function of the operating voltage is shown in fig. 7, curve A. It is seen that the slope is very slight (~ 0.005 per cent per volt), which is considerably less than that of a good Geiger tube. Thus we have essentially complete detection of the beta-rays of tritium, with energy between, say, 30 ev. and 17 kev., although the tube is operating throughout the region in a proportional manner. The upper limit of the plateau is due

to extraneous effects leading to spurious pulses, *e. g.*, the release of photoelectrons from the gas or wall. It must be stressed that this tube was not designed to perform as a Geiger counter; nevertheless it offers the same advantages of quantitative detection as a Geiger type and the long plateau means that no difficulties arise regarding voltage stability.

A very interesting result is shown in fig. 7, curve B. This curve refers to the same tube as for curve A of fig. 7, but in this instance it was filled with commercial nitrogen and the fluorescence X-rays of copper were passed into the tube through a thin window. Again, a plateau of nearly zero slope is obtained with the tube operating in the proportional

Fig. 8.



Variation of gas multiplication with voltage on counter for argon and methane and for nitrogen.

region. Due to the fact that most of the photoelectrons were released at the copper wall of the counter they had an average energy considerably less than 8 kev. We can confidently deduce that if radiation gives rise to 100 ion pairs in the nitrogen complete detection can be achieved as in a Geiger tube. For instance, at a gas gain of 100 the counter is stable and the pulses are well above the peak noise level.

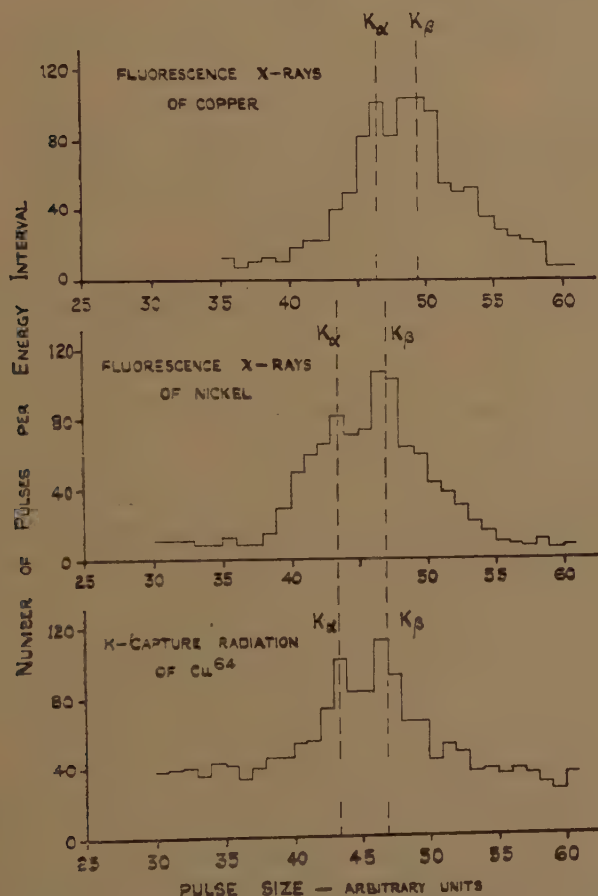
The importance of the result for nitrogen lies in the fact that such an instrument can replace a Geiger tube in many applications. Thus fast beta-particles or cosmic rays giving the minimum specific ionization will produce in a path length of 2 cm. sufficient ion pairs to ensure complete detection. The proportional tube used in this way with nitrogen as the

gas is free from all the difficulties associated with the finite life of the conventional Geiger tubes filled with a mixture of argon and organic vapour.

Rate of Variation of Amplification with Voltage.

It is important that the rate of change of gas gain with voltage should not be excessive. Two examples of results obtained with typical tubes are given in fig. 8. In view of the wide application of argon we have measured its absolute gain. The interesting case of nitrogen is given in terms of relative gain. Both are straight lines on the logarithmic scale as expected (Korff 1945, Corson and Wilson 1948) and the low values of the gradients are very satisfactory.

Fig. 9.



Comparison of K-capture radiation with fluorescence X-radiation

C. Identification of Radiation.

As an example of the application of this method to the study of K-capture phenomena curves are shown in fig. 9 for the K X-rays of copper and nickel, together with the curve obtained for the electromagnetic

radiation emitted by radioactive copper. This source decays by the three processes :

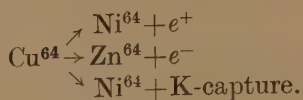
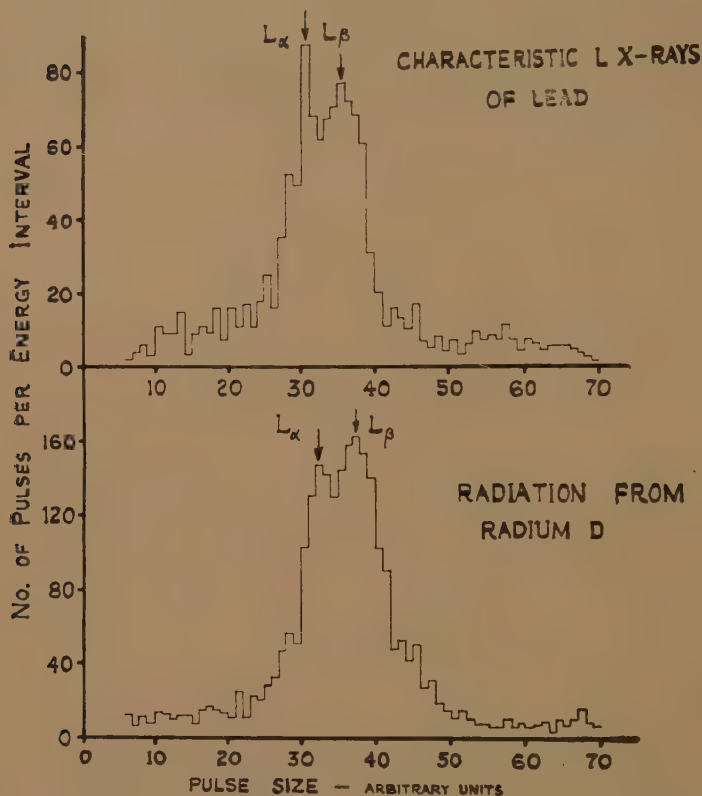


Fig. 10.



Identification of L X-rays emitted by RaD.

It was covered with polythene of sufficient thickness to stop all the electrons and positrons; the radiations which will be detected are therefore :

- (a) K X-rays resulting from the K-capture,
 - (b) annihilation quanta of energy 0.51 Mev.,
 - (c) gamma quanta of 1.3 Mev. energy
- and (d) the natural background of the tube.

A very brief analysis shows that the X-radiation appears quite distinctly above the heterogeneous background and corresponds exactly with the K-radiation of nickel. The resolving power of the instrument is quite adequate to separate it from the characteristic radiation of copper, which differs by only one unit of atomic charge. L-radiation

can likewise be identified ; thus the L X-rays emitted in the disintegration of radium D can be shown to coincide in energy with those of bismuth (see fig. 10) and to differ from those of lead, indicating their origin in the product nucleus radium E.

D. Resolving Power.

The theoretical analysis of the resolution of the instrument presents very considerable difficulties. Thus we have to consider :

(a) the statistical fluctuation of the number of ion pairs formed in the gas by the photoelectrons or beta-rays. All of the energy of the particles is absorbed in ionization, excitation and dissociation processes. The probability that an electron forms an ion pair is therefore an involved function of the energy, and this function will be different for different gases. It cannot be assumed that if n is the average number of ion pairs released by an electron of energy E the probable variations of n will be $\pm\sqrt{n}$.

(b) Statistical fluctuation in gas amplification must enter into the fluctuation in pulse size. This fluctuation is probably a sensitive function of the nature of the gas and the gas gain A employed. It will depend also on the value of n since this determines the total number of electrons collected.

Up to the present we have proceeded in an entirely empirical manner and we conclude from the examination of the various histograms in argon-methane that the full width at half amplitude of a distribution for a homogeneous radiation giving rise to A ion pairs on the average may be expressed approximately as $A \propto \sqrt{n}$, where α has a value between 1 and 1.5, and generally rather closer to the former.

It should perhaps be noted that the width obtained experimentally includes the fluctuations due to the noise level of the amplifier. It is hoped that we can investigate the rôle of the various factors in more detail in later work, especially in view of the fact that the peak width measured here is considerably less than the value of $2\sqrt{2n}$. A at 60 per cent of full amplitude, although this latter value is to be expected from elementary considerations of the normal statistical fluctuations in the initial ionization and subsequent gas amplification.

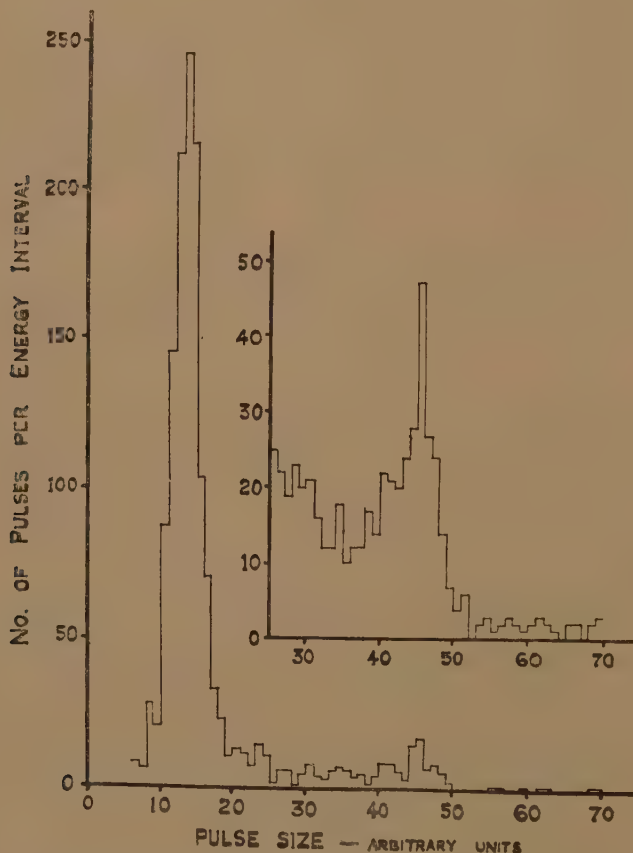
Application to the Gamma-radiation of Radium D.

The gamma rays of radium D have been the subject of a considerable number of investigations (Frilley, Surugue, and Tsien San-Tsiang 1946, Tsien San-Tsiang, and Marty 1946, Tsien San-Tsiang, 1944, and Stahel 1938). It was thought to be a convenient source with which to test the possibilities of the technique discussed here. The sources used were sometimes separated from the daughter elements, radium E and radium F, and occasionally an equilibrium mixture of all three was used. In the latter case sufficient polythene to absorb all the beta-particles of radium E was interposed between the source and detector.

The results of an investigation of the X-rays have been shown in fig. 10, above. It is seen that the radiation corresponds to the L X-rays of radium E (bismuth).

In fig. 11 the histogram shows the large peak due to the X-radiation and the small peak due to the unconverted gamma ray of energy 46.7 kev. (Frilley, Surugue, and Tsien San-Tsiang 1946); our measurements give the quantum energy as 46.0 kev. The latter is shown in

Fig. 11.



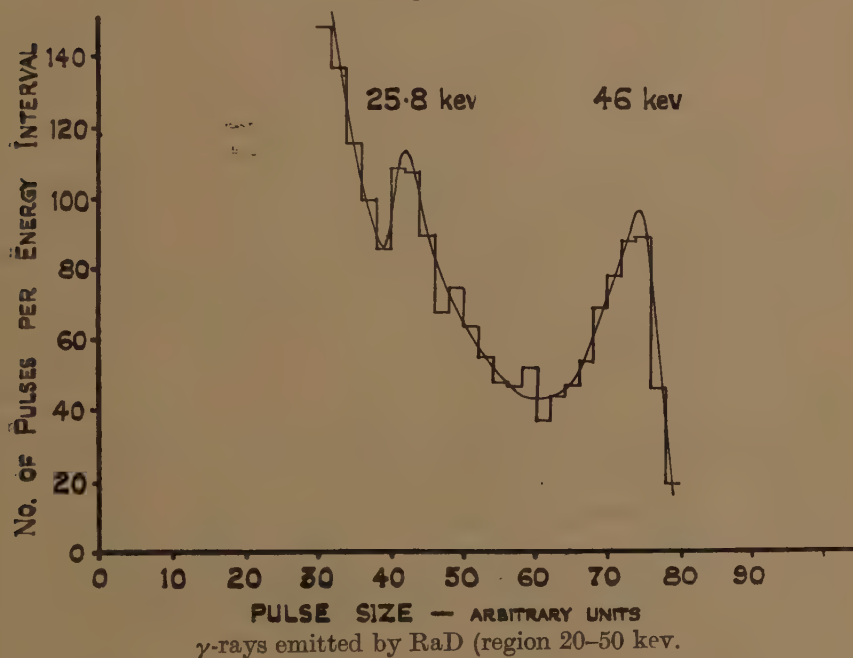
L X-rays of RaD and γ -radiation of energy 46 kev.

detail in the inset figure. The relative intensities of the X-radiation and this gamma ray were obtained by evaluating the area under the peak in the inset figure and multiplying the area under the X-ray peak of the main figure by the ratio of the lengths of film analysed. This gave the ratio of the areas as 18 : 1. Allowing for the effect of the absorbers between the source and the gas and for the efficiencies of detection this indicates an intensity ratio of 11 : 1.

In fig. 12 our analysis reveals the presence of a gamma-ray of energy 25.8 kev., presumably corresponding to the known gamma-ray of energy

23.4 kev. (Frlley, Surugue, and Tsien San-Tsiang 1946); our measurements show that the relative intensities of this line and the line of energy 46.0 kev. are in the ratio of 1 : 8. This is not in agreement with previous work which indicates that this ratio is approximately 1 : 3 (Frlley, Surugue, and Tsien San-Tsiang 1946). Assuming that 3 quanta of the harder radiation are emitted per 100 disintegrations (Frlley, Surugue, and Tsien San-Tsiang 1946), the radiation of energy 25.8 kev. is emitted with an intensity of 0.4 quanta per 100 disintegrations.

Fig. 12.



The histogram of fig. 13 indicates clearly the presence of a soft gamma-ray of energy 7.8 kev. Here we estimate (after correcting for absorption in air and the mica window of the counter) that the intensities of the quanta and L X-rays are in the ratio 0.22 : 1.

The intensities may be tabulated as follows :—

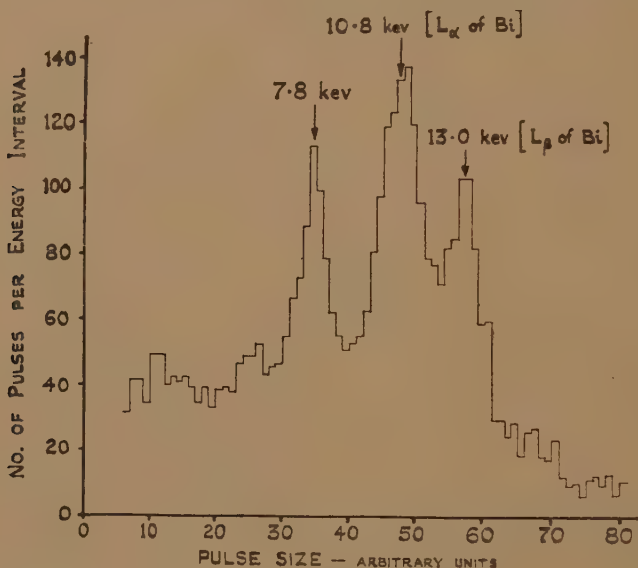
Energy-kev.	7.8	10-15 (L X-rays)	25.8	46.0
Rel. intensity	7.3	33	0.4	3.0

It must be stressed that all the data given here for radium D was obtained with very weak sources (\sim one microcurie) and exposure times of some minutes. This contrasts very favourably with the difficulties associated with the spectrographic method which involves relatively

strong sources and long exposure times. All of the gamma-ray analysis was done with a counter chamber of diameter 7 cm., filled with argon and methane (60 cm. and 15 cm. respectively).

It seems from our histograms that there may be a number of gamma-rays of other energies of weak intensity, but we have not chosen at this stage to investigate such possible radiations in detail.

Fig. 13.



Soft γ -radiation and L X-rays from RaD. Source was Ra(D+E+F) in magnetic field at 8 cm. from mica window (1.8×10^{-3} cm. thick).

We are pleased to thank Prof. P. I. Dee for his help and advice throughout this work. We are indebted to Mr. J. T. Lloyd for his very valuable assistance in the design and preparation of the counters. We thank Miss C. F. Lees for assistance in counting.

REFERENCES.

- RUTHERFORD, E., and GEIGER, H., 1908, *Proc. Roy. Soc. A*, **81**, 141.
 KORFF, S. A., 1946, *Electron and Nuclear Counters*, Chap. 3 (Van Nostrand).
 CORSON, D. R., and WILSON, R. R., 1948, *Rev. Sci. Inst.*, **19**, 207.
 GREISON, K., and NERESON, N., 1942, *Phys. Rev.*, **62**, 316.
 COMPTON, A. H., and ALLISON, S. K., 1936, *X-Rays* (Macmillan).
 GRAY, L. H., 1944, *Proc. Camb. Phil. Soc.*, **40**, 72.
 EISEL, A., 1929, *Ann. Phys. Leipzig*, **3**, 277.
 GERBES, W., 1935, *Ann. d. Phys.*, **23**, 648.
 SMYTH, H. D., 1931, *Rev. Mod. Phys.*, **3**, 347.
 FRILLEY, M., SURUGUE, J., and TSIEN SAN-TSIANG, 1946, *J. de Phys. et le Rad.*, ser. 8, **17**, 350.
 TSIEN SAN-TSIANG and MARTY, C., 1945, *Comptes Rendus*, **220**, 688.
 TSIEN SAN-TSIANG, 1944, *Comptes Rendus*, **218**, 503.
 STAHEL, E., 1938, *Helv. Phys. Acta*, **8**, 651.
 RICHARDSON, H. O. W., and LEIGH-SMITH, A., 1937 *Proc. Roy. Soc. A*, **160**, 454.

III. *Investigation of Soft Radiations—II. The Beta Spectrum of Tritium* *.

By S. C. CURRAN, J. ANGUS, and A. L. COCKROFT,

Department of Natural Philosophy, University of Glasgow †.

[Received October 12, 1948.]

SUMMARY.

The β -spectrum of tritium is investigated between 0.5 kev. and the upper energy limit, 17.9 kev. Detailed results near the low energy end and close to the upper limit are presented. The spectrum is shown to be simple and in reasonable agreement with Fermi theory. There is evidence of some deficiency in β -rays of low energy. The best fit with theory is obtained for a neutrino mass $\mu = m/500$ (m = electron rest mass).

INTRODUCTION.

THE proportional counter technique discussed in the preceding paper (Part I.) is particularly well suited to the investigation of the beta-spectra of radioelements of low limiting energy. There are two chief difficulties associated with the accurate analysis by normal spectrographic methods of a spectrum at the low energy end :

(a) the finite thickness of source and support introduces uncertainties in the straggling and reflexion of the electrons ;

(b) the absorption on the thin window of the detector (normally a Geiger tube, or tubes) makes correction of the data at low energies a somewhat empirical procedure.

These two disadvantages can be overcome by introducing the radioactive element into the proportional counter as a gaseous constituent, and by using the counter to detect and measure the energy of the particles. The histogram of the pulses gives the form of the spectrum.

We have previously reported the preliminary results (Curran, Angus, and Cockroft 1948) obtained by this method for the nucleus H^3 . The following account extends the scope of the results.

The beta decay of H^3 is an allowed transition. Recent work (Novick, and Goldblatt 1947) on the determination of the half-life shows that it is probably between 10 and 12 years. On the other hand, a number of

* The subject matter of this paper was reported briefly by Prof. P. I. Dee at the Birmingham Conference on Physics (September 1948).

† Communicated by the Authors.

attempts (Watts, and Williams 1946) to determine the upper energy limit of the spectrum have given values ranging between 11 and 18 kev. Any of these values indicates that the transition is definitely allowed, but Konopinski (1947) has pointed out that there exists a theoretical discrepancy between the previously determined values of the energy limit E_0 and the half-life, the discrepancy being more serious for the lower experimental results.

The decay of H^3 is also of particular interest in view of the relatively simple structure of the nucleus and the possibility of determining the mass of the neutrino by examination of the shape of the spectrum near the upper energy limit. The low value of the limiting energy makes the shape a particularly sensitive function of the neutrino mass, provided that Fermi theory is assumed to hold.

APPARATUS AND RESULTS.

The technique of construction and application of the proportional tubes has been described in Part I. For earlier work on H^3 the counters were made with glass envelopes and aluminium cathodes. The data presented here was obtained with a long copper counter, inner diameter 2.25 in., effective length of 16 in., and a wire diameter of 4 mil. The end-effect of the tube had been carefully examined (see left-hand side of fig. 2 of Part I.), and the gas gain checked as constant over the extended region beyond the vicinity of the ends. This was done in order to correct for the distortion of the spectrum due to reduction of gain near the ends. The procedure adopted was as follows. A theoretical Fermi distribution for H^3 of upper energy limit E_0 equal to 18.0 kev. was evaluated, using the formula

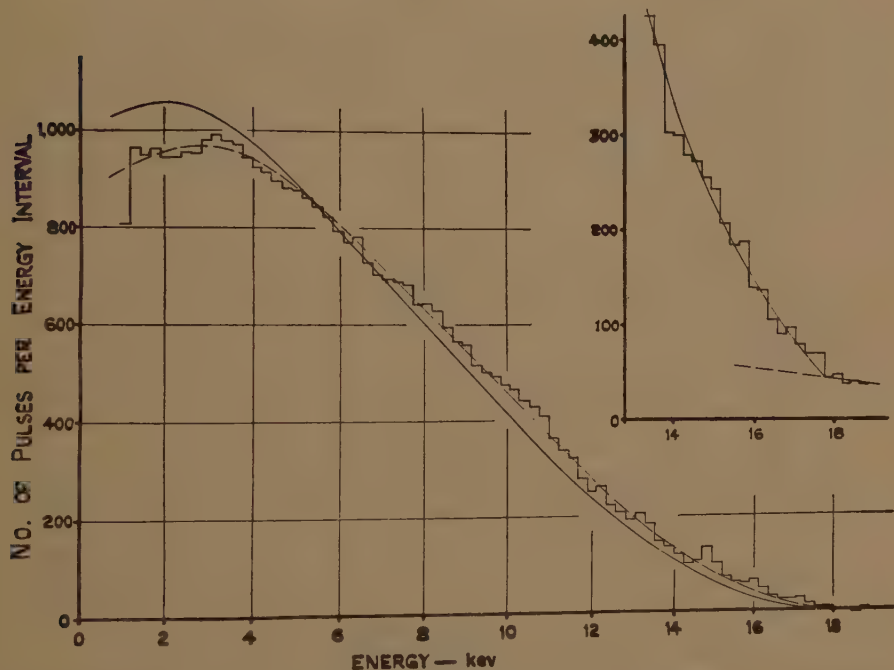
$$S(E) = (E_0 - E)^2 (1 - e^{-Z(0.5369/E)^{1/2}})^{-1} \cdot D, \quad . \quad . \quad . \quad (1)$$

where E is the energy, non-relativistic, Z is the nuclear charge (value 2) and D is the Dirac factor, of importance only for heavier nuclei. Near the ends of the wire the counter was considered to consist of small slices, in any one of which the gas gain was known as a fraction of the gain in the main tube. In this way the form that the Fermi distribution would take in each of the slices could be calculated. The relative intensity of each such distribution was known from the volume of the slice as a fraction of the total volume of the truly effective part of the counter. Adding such Fermi distributions to the main Fermi curve for E_0 equal to 18 kev. gave the Fermi curve as modified by the counter. The total modification was found to be very slight and certain to lead to no appreciable error, except possibly at very low energies. Above the point of maximum intensity on the curve the change in shape was negligible, while below the maximum the intensity was increased on the average by about 5 per cent. In view of the relatively small modification ascribable to end-effect we have chosen to show only this modified theoretical Fermi distribution as the full curve in fig. 1.

It may be noted that a second and even more accurate method of correcting for end-distortion is quite practicable. If the counter wire is divided into two unequal portions by a glass bead the difference spectrum obtained by subtracting the distribution observed in the shorter from the distribution observed in the longer will be automatically correct.

The histogram of the beta-pulses is shown in fig. 1 and the modified Fermi curve is adjusted so that the area under it is the same as the area under the smooth dashed curve through the experimental values. The energy values were obtained as previously (Part I.) by observing the

Fig. 1.



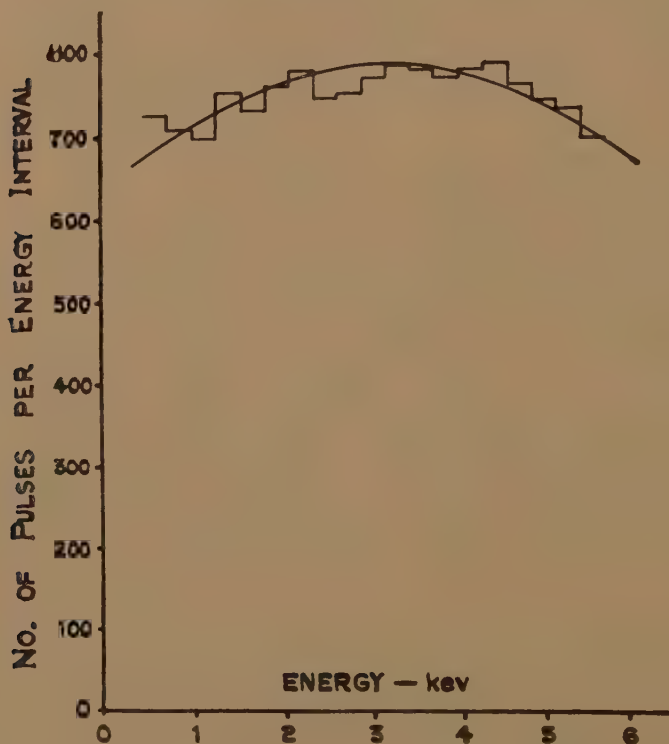
β -spectrum of H^3 . The slightly modified theoretical spectrum is indicated by the full curve.

pulses produced by the fluorescence X-rays of copper and nickel. The form of the curve near the end-point is shown in more detail in the inset diagram of the figure. In fig. 2 we show the results obtained in investigating the spectrum more thoroughly at low energies. Additional accuracy was obtained by increasing the gas amplification so that more detail could be observed in the softer region.

The marked step at the beginning of the distribution in fig. 1 is thus shown to be due to experimental limitations only and the curve through the values in fig. 2 would appear to be smooth and tending to decrease slowly with decreasing energy of the beta-rays. The maximum is somewhat wider and flatter than that predicted by theory. Moreover, the intensity at lower energies is a little less than theoretical and a little

greater than theoretical at higher energies. It seems rather unlikely that this slight disagreement with theory can be ascribed to any experimental limitation of the technique. We have established previously that the number of ion pairs formed in the gas mixture (argon and methane) is almost exactly proportional to the energy of the electron producing them. We have checked that the resolving power of the counter would not give rise to an effect of this kind of appreciable magnitude.

Fig. 2.



Detail of spectrum in the low energy region.

The form of the experimental curve would be modified considerably if the tritium was absorbed in the wall material to an appreciable extent. A number of checks proved conclusively that the tritium, which was released from palladium by heating, mixed with an excess of hydrogen and then passed into the counter, existed entirely in the gaseous state :

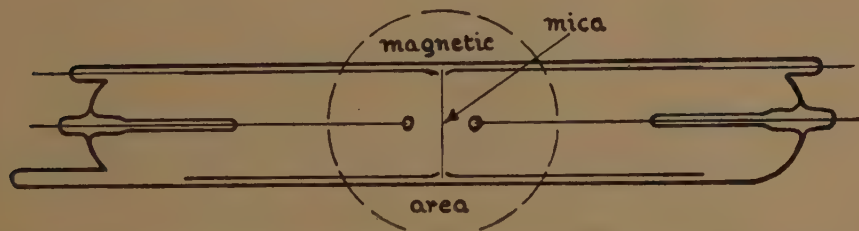
(a) the tritium could be pumped out of the counter readily and completely, and the natural background counting rate restored ;

(b) the rates of counting in two counters filled with the same mixture were accurately proportional to counting volume and not to the surface areas of the tubes ;

(c) when the gaseous mixture in a counter was shared with an exhausted counter the activities as measured in both were very accurately proportional to the volumes. This sharing experiment was checked to an accuracy better than 1 per cent.

The possibility that the spectrum was complex was investigated. The chemical procedure for preparation of the source was such as to make it extremely unlikely that any radioactive contaminant was present. The long life of the radioelement was an additional guide here. No gamma radiation of H^3 has been reported and if beta-rays of a softer spectrum were emitted in 20 per cent or more of the disintegrations, the complexity would be detected as a discontinuity in the histogram. Over most of the range of energy involved the quantum efficiency is greater than 25 per cent, and a discontinuity of 5 per cent or more would result. Nevertheless a direct and more sensitive search for associated quanta was made. For efficiency the source of H^3 was again gaseous, and the arrangement adopted is shown in fig. 3. Two Geiger tubes

Fig. 3.



Coincidence counting arrangement used to establish simplicity of β -disintegration process.

were enclosed in the same envelope and separated by a thin mica foil (1 mg/cm^2) which served two purposes :

(a) preventing interaction of discharges between the two counters,

(b) absorbing beta-particles of H^3 which might trigger both tubes. The filling mixture consisted of argon (74 cm.) and alcohol (1.5 cm.) to which hydrogen and H^3 (total pressure 0.50 cm. Hg) was added. A magnetic field was used to eliminate some natural coincidences, though this was not essential.

With H^3 in the vessel no increase in the rate of coincidence counting was detected. This test established that less than 1 per cent of the disintegrations of H^3 could be complex with gamma radiation following beta emission.

Background coincidence rate $= (2.50 \pm 0.16)$ per min.

Coincidence rate with H^3 in vessel $= (2.25 \pm 0.10)$ per min.

Increase in counting rate due to $H^3 = 880$ per min.

Increase in coincidence rate due to $H^3 = (-0.25 \pm 0.26)$ per min.

It was estimated that each counter would be traversed by 10 per cent of any radiation leaving the other as source and the average efficiency of detection of such radiation is ~ 10 per cent. Thus in the above experiment a maximum increase in the coincidence rate of about 8.8 per min. is possible for complete beta-gamma correlation. It is safe to conclude that less than 1 per cent of the disintegrations are complex since the measured maximum probable increase is 0.01 per min.

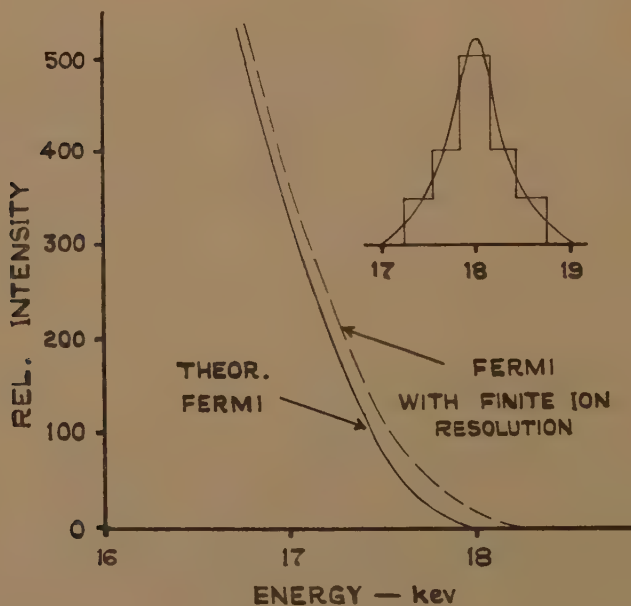
We have considered the possibility that nuclear screening might affect the shape of the Fermi distribution. However, the calculated curve for $Z=1.7$ differs little from that for $Z=2$ given here.

EFFECT OF RESOLVING POWER ON THE SPECTRUM.

1. Upper Energy Limit.

We have already noted that the finite resolving power of the tube does not appreciably affect the shape of most of the beta-distribution.

Fig. 4.



Modification of Fermi distribution near end-point due to finite resolution.

This is largely because of the slowly varying nature of the curve. Near the end-point, however, significant modification is introduced. Thus, suppose that a homogenous radiation of energy 18 keV. gives an experimental distribution like that shown in the inset of fig. 4. This histogram is a reasonable approximation to practice (see Paper I.) since it has a full width at half-amplitude of about $1/\sqrt{n}$ of the energy of the radiation. In this case the true Fermi distribution of fig. 4 (solid curve) with end point E_0 at 18 keV. would be observed experimentally as the dashed curve. We see that only in the immediate vicinity of the end-point,

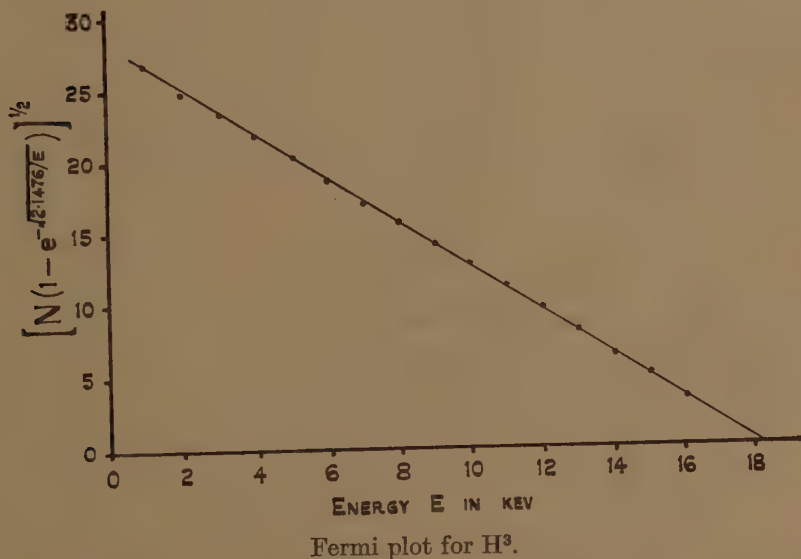
i. e., within 1 kev. of E_0 , is the distribution significantly altered by the experimental resolving power. The finite resolution gives rise to a tail of very small intensity extending beyond the real end-point. The effect makes it somewhat more difficult to determine with very high precision the true value of E_0 .

Considering now the inset curve of fig. 1 we estimate that the value of E_0 for H^3 is given by

$$E_0 = 17.9 \pm 0.3 \text{ kev.}$$

This value is consistent with a number of independent determinations made by us and it is considered to be more accurate than our own previous value of 16.9 ± 0.3 kev. (Curran, Angus, and Cockroft 1948). It is in better theoretical agreement (Pruett 1947) with the measured half-life of 10 to 12 yr. than the recently published value of 11 kev. In fig. 5

Fig. 5.



we have plotted $[N(1 - e^{-\sqrt{2 \cdot 1476/E}})]^{1/2}$ against E , where N is the intensity as given by the smooth curve through the experimental histogram of fig. 1. For a spectrum agreeing with Fermi theory this plot should result in a straight line. The points lie on or close to a straight line, and this line intercepts the E -axis at a value of $E_0 = 18.1$ kev.

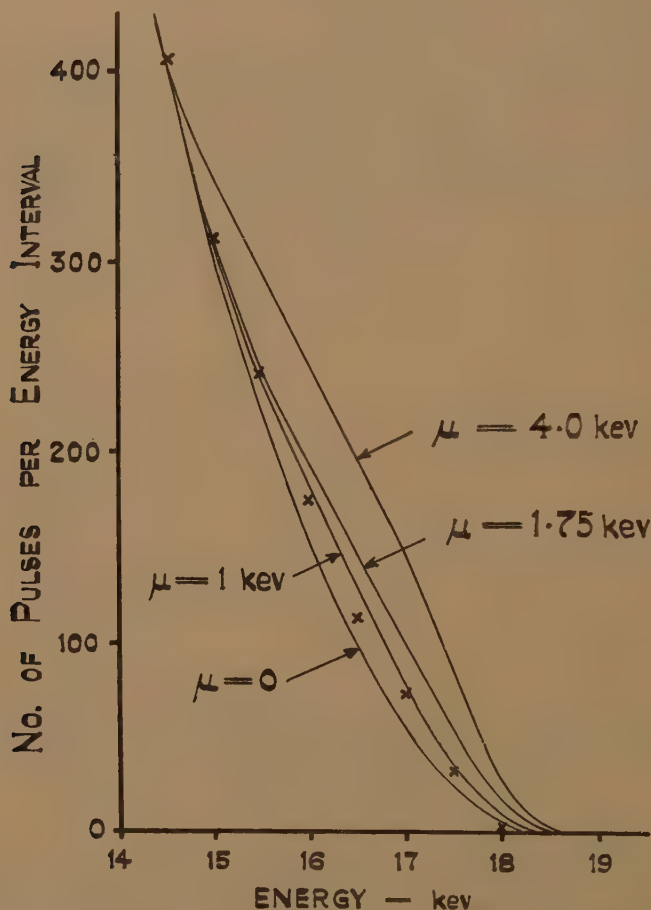
2. Neutrino Mass.

In fig. 6 we show a set of theoretical curves corresponding to neutrino rest mass (μ) values of zero, $m/500$, $m/300$ and $m/125$. The curves have been modified slightly to allow for the resolution of the apparatus (inset, fig. 4), but the general shape remains unaltered. It is found that the smooth curve through our results (inset, fig. 1) agrees closely with the curve for $\mu = m/500$. The experimental values seem to be inconsistent with a zero mass. Additional accuracy of the data is perhaps necessary

to establish with certainty such a conclusion, but assuming the validity of the neutrino theory we can at least deduce from the results that μ is given by the equation :

$$0 \leq \mu < m/300.$$

Fig. 6.



Evaluation of neutrino mass. Experimental values are indicated by crosses.

We are indebted to Professor P. I. Dee for his advice and help throughout the work, and grateful to Mr. B. Touschek for the theoretical discussions which we enjoyed with him. We thank Miss C. F. Lees for assistance in counting.

REFERENCES.

- CURRAN, S. C., ANGUS, J., and COCKROFT, A. L., 1948, *Nature, Lond.*, **162**, 302.
 NOVICK, A., 1947, *Phys. Rev.*, **72**, 972; GOLDBLATT, M., 1947, *Ibid.*, **72**, 973.
 WATTS, R. J., and WILLIAMS, D., 1946, *Phys. Rev.*, **70**, 640 and references cited.
 KONOPINSKI, E. J., 1947, *Phys. Rev.*, **72**, 518.
 PRUETT, J. R., 1947, *Bull. Am. Phys. Soc. A*, **22**, 5.

IV. *A Contribution to the Theory of Liquid Helium II.*

By N. F. MOTT, F.R.S.,

H. H. Wills Physics Laboratory, University of Bristol *.

[Received August 26, 1948.]

SUMMARY.

A THEORY is given of certain properties of liquid helium, in particular

- (a) Its behaviour as a mixture of a superfluid and a normal fluid.
- (b) The behaviour of the Rollin film.
- (c) The existence of a critical velocity in flow through capillaries above which viscosity appears.

The theory is based on London's concept of a condensed Einstein-Bose gas and Bijl, de Boer and Michels' explanation of the energy gap between the lowest state and the excited states.

It seems possible to show that, to the approximation to which interaction between pairs of atoms alone is considered, the theory accounts well for the phenomena that depend on the "two fluid" model of Landau and Tisza.

1. INTRODUCTION.

The purpose of this paper is to give in outline a theory of some properties of liquid helium II, in particular of its behaviour in capillaries. The theory is based on London's concept of a condensed Einstein-Bose gas and on certain ideas put forward by Bijl, de Boer and Michels (1941).

The following phenomenological description of liquid helium seems to be in accord with the experimental facts :

(a) *Properties of helium at the absolute zero of temperature.*

In the theory of solids (*cf.* for example Mott and Jones 1936, p. 16), it is usual to plot the total energy of a solid at the absolute zero against volume. If a complete theory of liquid and solid helium existed, it should be possible to do this for helium. A schematic representation of the sort of curve to be expected at the absolute zero of temperature is shown in fig. 1. For solid helium one would get a curve of the normal type with a minimum, and for the superfluid a curve with a lower minimum (A). Then the corresponding volume V_A will represent the volume of the superfluid under zero pressure, for values of the volume between V_B and V_C solid and liquid will exist in equilibrium, and the slope of the tangent BC gives the pressure under which solidification takes place. The only parts of the curve which can be investigated experimentally are from A to B and C to D and smaller volumes.

* Communicated by the Author.

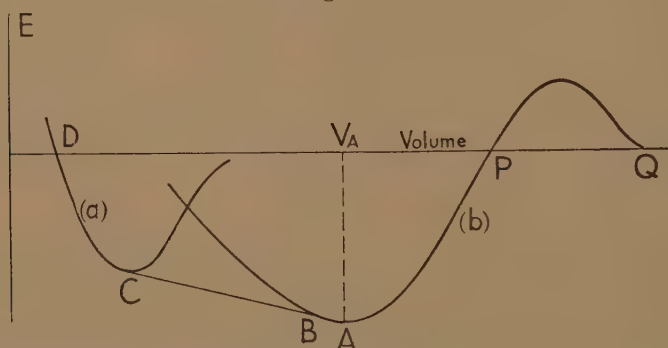
We believe it possible that calculations for other substances obeying Einstein-Bose statistics might show a similar behaviour (*e. g.* H_2); but the superfluid will never be observed unless the minimum A is below C. It is in this that helium is probably unique.

(b) *Properties of liquid helium at temperatures below the transition point (2.19°K.).*

A number of experimental facts (second sound, fountain effect, vanishing viscosity in capillaries, etc.) have been interpreted in terms of the following model (Landau 1941, Tisza 1947, Dingle 1948): Liquid helium II behaves as if it were a mixture of two fluids; these are

- (i) The superfluid, which cannot carry Debye waves (phonons^{*}), and must be regarded as a system in its lowest quantum state.
- (ii) The normal fluid, which has entropy and in fact behaves, both as regards its heat content and viscosity, like a normal liquid or gas.

Fig. 1.



Energy E of helium at absolute zero plotted against volume (schematic).
Curve (a), solid; curve (b), superfluid.

At any temperature T the relative amounts of normal and superfluid present in thermodynamic equilibrium are determined[†], and if the temperature is changed rapid transformation of one into the other occurs. But the essential point in the model is that if the normal fluid and superfluid are set in motion relative to each other, *there is no transfer of momentum from one to the other*. This assumption is implicit in the theoretical treatments of second sound given by Landau, Tisza and Dingle, and in the experimental determination of the ratio of the densities of the two fictitious fluids given by Andronikashvili (1946). As we shall show, the behaviour of helium in Rollin films and capillaries supports the assumption. The absence of momentum transfer appears, however,

^{*} In Tisza's version of the theory this assumption is not made; it seems to us to be vital, as will be shown below.

[†] The ratio as a function of T can be determined experimentally in a number of ways, of which that of Andronikashvili (1946) is perhaps the most direct.

to depend on the relative velocity being lower than a critical value v_c , which depends on the diameter d of the capillary or on the thickness d of the film in which the helium is flowing; according to the empirical relation of Bijl, de Boer and Michels (1941), v_c is given by

$$v_c \sim h/md,$$

where m is the mass of the helium atom. This relation is valid only as regards its order of magnitude; a linear dependence on $1/d$ is not yet proved.

In § 4 we shall give a theoretical basis for the phenomenological model. This will be based on the conception of Einstein-Bose condensation, and will extend the ideas originally due to London.

2. THE SUPERFLUID AT THE ABSOLUTE ZERO.

In this section we consider the form of the wave function of a superfluid at the absolute zero of temperature. This, of course, depends on the form of the interaction between two helium atoms. We describe this by a function $V(r)$ which denotes the potential energy of a pair of helium atoms at a distance r apart. $V(r)$ consists of the sum of two terms:

- (a) A weak Van der Waals attractive term $-C/r^6$.
- (b) A strong repulsive term; for the purposes of this paper we shall treat this by representing the helium atoms as impenetrable spheres of radius b .

We shall make the further assumption that the attractive term is so weak that no stable helium molecule can be formed for helium atoms in their lowest electronic state. From this assumption it follows that, if $\psi(\mathbf{r}_1, \mathbf{r}_2)$ is any function of the co-ordinates $\mathbf{r}_1, \mathbf{r}_2$, the energy integral

$$\iint \psi^* \left\{ -\frac{\hbar^2}{2m} (\nabla_1^2 + \nabla_2^2) + V(|\mathbf{r}_1 - \mathbf{r}_2|) \right\} \psi d\tau_1 d\tau_2 \quad . \quad . \quad (1)$$

is always positive.

We turn now to the wave function of an assembly of N helium atoms in its normal state. In the absence of interaction we could write this as a simple product:

$$\Psi_0(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N) = \psi(\mathbf{r}_1) \psi(\mathbf{r}_2) \dots \psi(\mathbf{r}_N),$$

where $\psi(\mathbf{r})$ is the wave function for a particle with co-ordinates $\mathbf{r} = (x, y, z)$ in its lowest state in the containing box. Thus if the box is a cube of side L ,

$$\psi(\mathbf{r}) = (8/L^3)^{1/2} \sin(\pi x/L) \sin(\pi y/L) \sin(\pi z/L) \quad . \quad . \quad . \quad (2)$$

With the hard sphere model of the helium atom, one cannot, of course, use perturbation theory to calculate the effect of interaction; this would lead to an infinite energy. We therefore introduce a modified wave function Ψ of the following form:

$$\Psi(\mathbf{r}_1, \dots, \mathbf{r}_N) = \Psi_0(\mathbf{r}_1, \dots, \mathbf{r}_N) \prod_{s,t} f(|\mathbf{r}_s - \mathbf{r}_t|), \quad . \quad . \quad . \quad (3)$$

where $f(r)$ is some function that vanishes when $r=2b$ and tends to unity when r tends to infinity. A suitable form is

$$f(r)=1-e^{-\beta(r-2b)}.$$

One can then attempt to calculate the energy, $W(\beta)$,

$$W(\beta)=\int \Psi^* \left\{ -\frac{\hbar^2}{2m} \sum_s \nabla_s^2 + \sum_{s,t} V(|\mathbf{r}_s - \mathbf{r}_t|) \right\} \Psi d\tau_1 \dots d\tau_N, \quad (4)$$

and choose β in such a way as to make $W(\beta)$ a minimum.

As long as the volume is great enough in comparison with Nb^3 for it to be possible to calculate the energy from two-body encounters only, with neglect of three-body encounters, it follows, from the assumption that (1) is positive, that (4) is positive too. Thus to this approximation no condensation can occur, and the total energy will be positive initially (as shown in fig. 1 for the region PQ). Attraction and consequent condensation must occur for smaller volumes when this approximation breaks down. We have not attempted to treat this case.

In a subsequent note Dingle has estimated the terms in the energy $\int \Psi^* \{ -(\hbar^2/2m) \sum_s \nabla_s^2 \} \Psi d\tau$ by the above method. It comes out to be, per atom,

$$W_{\text{kin}} = 15\hbar^2 b / ma^3, \quad (5)$$

where a^3 is the atomic volume L^3/N .

3. THE THICKNESS OF THE ROLLIN FILM AT THE ABSOLUTE ZERO OF TEMPERATURE.

Bijl, de Boer and Michels (1941) have suggested the following explanation of the formation of a thick film of helium II on any solid surface in contact with the liquid. It is assumed that the cohesive energy between a layer of helium and any solid in contact with it is greater than that between any two layers of helium. This would follow from the accepted theory of Van der Waals forces. This would ensure the formation of a thin, or monomolecular film. The question is, why does a thick film form? The above authors, taking as a model for helium II an Einstein-Bose gas without interaction between atoms, state correctly that the zero-point energy of each atom is $\hbar^2/8md^2$ for a film of thickness d . For unit area of film at height x above the surface of the liquid the potential energy is thus

$$\left\{ \frac{\hbar^2}{8md} + mgxd \right\} / a^3, \quad (6)$$

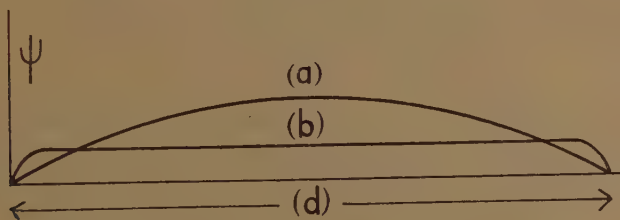
where $1/a^3$ as before is the number of atoms per unit volume. Minimizing this with respect to d , one finds

$$d = \hbar / 2m \sqrt{(2gx)}.$$

When the effect of interaction between atoms is introduced into the theory, we do not believe this explanation to be tenable. The zero-point energy $\hbar^2/8md^2$ depends on the assumption that the wave function is of the form $\psi = \sin(\pi x/d)$. But such a wave function would give a very non-uniform distribution of the fluid's density across the film. This distribution could not possibly correspond to the state of lowest energy. If one were to calculate the total energy (including attractive forces) from a wave function such as (3), and if one varied ψ as well as f so as to obtain a minimum energy, it is obvious that a form for ψ , such as that of curve (b) in fig. 2, would be obtained instead of the sine function (2), curve (a), flat except at the edges. Such a wave function would give no energy term proportional to $1/d^2$, but a surface energy independent of the thickness.

A more likely explanation of the Rollin film seems to us to be that due to Frenkel (1940) and Schiff (1941), namely, that it is simply due to the long-range Van der Waals forces between helium and the substrate.

Fig. 2.



Wave functions for helium atoms in thin film.

(a) According to zero order approximation.

(b) True form.

These authors show by an argument similar to that given above that on this assumption a *stationary* film would have thickness at height x

$$d \propto 1/x^{1/3}.$$

This explanation of the origin of the film is in accordance with recent observations of Kisternacher (1947), showing that at temperatures at which liquid helium I is stable, thick films condense from the vapour on a solid surface in equilibrium. One might expect similar results to occur with hydrogen. The abnormal property of films of helium II is in our view their high velocity of creep, *i. e.* their vanishing viscosity, rather than their thickness.

4. HELIUM II AT A FINITE TEMPERATURE.

As first emphasized by London (1938) in this connection, it is characteristic of the Einstein-Bose statistics, applied to an ideal gas with a constant number of molecules, that below a certain critical temperature T_c of the order \hbar^2/ma^2k , a finite proportion of the molecules will be found in

the ground state. Thus for $T < T_c$ the number in the ground state is proportional for constant density to the total number N of molecules present, while the number in each excited state of the continuum is independent of N .

The Einstein-Bose statistics gives for the specific heat at low temperatures

$$C = \text{const. } T^{3/2}.$$

The observed specific heat increases with a much higher power of the temperature than this (empirically as about T^5); it has been pointed out by many authors that this could be explained if there were an energy gap between the lowest state of the Einstein-Bose continuum and the first excited states. The paper by Bijl, de Boer and Michels (1941) already quoted, explains how this gap can arise. As their explanation is essential for our discussion of helium films, it will be recapitulated here.

Consider two helium atoms, described by wave functions $\exp(ikx)$, $\exp(ik'x)$. Then the symmetrical function describing the two of them will be $\Psi = 2^{-1/2} \nu \exp(ikx_1) \exp(ik'x_2) + \exp(ik'x_2) \exp(ikx_1)$. ν is here a normalizing factor. The probability density $|\Psi|^2$ is thus

$$|\Psi|^2 = \nu^2 [1 + \cos \{(k-k')(x_1-x_2)\}].$$

If $k \neq k'$, then ν^2 will be equal to $1/V$, where V is the volume in which the two particles move. This follows from the fact that the mean value of the function $\cos \{(k-k')(x_1-x_2)\}$ may be taken to be zero. But if $k = k'$, the cosine is unity, so ν^2 must be taken to be $1/2V^2$; in this case then

$$|\Psi|^2 = 1/V^2.$$

It follows that, when the two particles are very close together ($|x_2 - x_1|$ small), $|\Psi|^2$ is twice as great when $k \neq k'$ as when $k = k'$. Thus any short-range interaction between the particles will contribute only half as much to the total energy when the particles are in the same state as when they are in different states. This result is quite independent of the use of perturbation theory, and would follow for an estimation of the interaction energy such as that of § 2; it depends of course on the restriction to two-body collisions.

The application to N particles is given in the paper quoted. Following the methods given there, and confining ourselves to two-body collisions, we obtain the following result: if we consider (a) a state of the whole system in which all particles are in the same state k , and (b) a state in which one particle is in the state k' , and all the others in states with the wave vector k , then, apart from the term

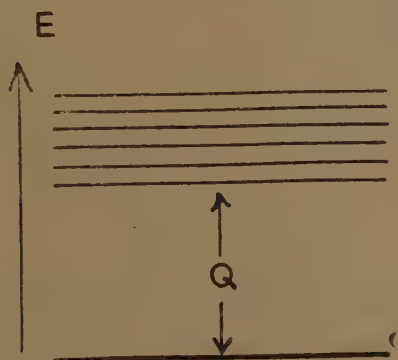
$$(\hbar^2/2m)(k'^2 - k^2),$$

the energy of the second term is higher than the first by the total interaction energy between a pair of particles, *i. e.* by a term of the type (5). The

energy spectrum of the assembly of N particles thus acquires an energy gap Q of the order kT_0 , where T_0 is several degrees and for given density independent of the volume, between the lowest state and the continuum of excited states. This is shown in fig. 3.

It will be noted that this result is only valid subject to condition (1)—i.e. that no molecule He_2 can be formed from atoms with no electronic excitation. Otherwise the analysis gives energies for the excited states lying below the normal state. This means that the approximation breaks down. This could have been predicted *a priori*, because if (1) is not valid the proper approximation of zero order would be an assembly of molecules He_2 . One cannot speak of an Einstein-Bose gas of particles attracting each other by central forces and capable of forming molecules.

Fig. 3.

Energy spectrum of system of N atoms with interaction.

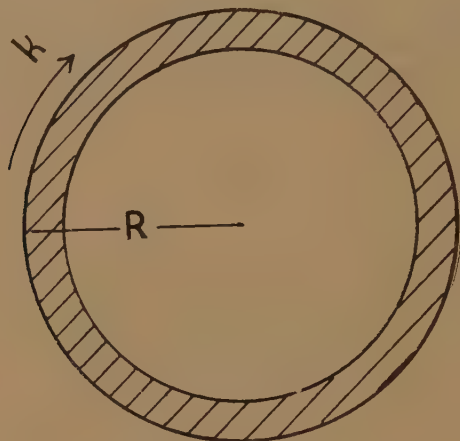
We have been unable to show when the two-body approximation breaks down and (we hope) leads to condensation, that nevertheless the excited states in this approximation are higher than the ground state. If this should turn out not to be the case, the theory of this paper would not be applicable to the actual superfluid, but only to an idealized model of hard spheres.

6. APPLICATION TO CAPILLARIES AND FILMS.

In terms of the two-fluid model, it is natural to identify the superfluid with the atoms in the lowest state and the normal fluid with the atoms in the excited states. The superfluid then makes no contribution to the entropy. It cannot transmit phonons. States of the whole system in which phonons (compressional waves) exist must be represented by states in the continuum of fig. 3, and thus above the energy gap. A theoretical prediction from the model is that the specific heat at low temperatures must vary according to an exponential term, say as $T^n e^{-Q/RT}$.

Our next task is to show that, if we set up a state in which the mean velocities per atom of the superfluid and normal fluid are unequal (by less than the critical velocity v_c), they will not change. For this purpose let us consider a circular tube at rest, containing helium II at temperature T which is flowing round it. This is shown in fig. 4. We define the state of the whole fluid by the assembly of wave numbers k which describe

Fig. 4.



Showing tube of rotating helium.

the wave functions e^{ikx} of the individual particles in the state of zero approximation. x is here a co-ordinate measured along the tube. Then all the atoms of the "superfluid" have the same wave vector k_0 , while those of the "normal fluid" will have a range of wave vectors k such that $\hbar^2 k^2 / 2m$ varies over a range comparable with kT , but owing to collisions with the walls the mean value of k for the normal particles is zero. The distribution function envisaged is shown in fig. 5 (a), where the number N

Fig. 5 (a).

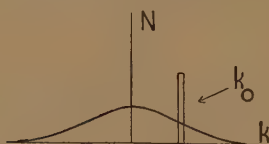
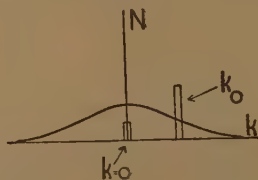


Fig. 5 (b).



of particles in each state is plotted against k . We envisage a constant interchange between normal and excited atoms; why then does not the superfluid slow down?

Slowing down could occur in two ways. The atoms could be transferred *one by one* from the state k_0 to a state of lower momentum, say the state $k=0$. The state in which for all atoms of the superfluid k is equal to zero has, of course, the lowest free energy. But, to reach it,

it is necessary to go through the intermediate state in which some of the condensed atoms are in states for which $k=0$ and some in states k_0 ; this intermediate state is shown in fig. 5 (b). It is clear that, since the (positive) interaction energy between the atoms occurs twice for atoms in different states, state 5 (b) will have higher free energy than state 5 (a). Thus the state 5 (b) is metastable for transitions of this type.

The second possibility is that all the atoms of the superfluid should jump at once from the state k_0 to a state with lower momentum. The momentum hk is, however, quantized; we may set, according to Bohr's equation for the angular momentum,

$$k=n/2\pi R,$$

where n is an integer. It follows that the value of k for each particle cannot change by less than $1/2\pi R$. In a transition of the type envisaged, therefore, the total change of momentum of the N atoms of the superfluid is

$$N\hbar/R.$$

Now if A is the area of the cross-section of the tube, and $1/a^3$ the number of atoms per unit volume, this is of order

$$\hbar A/a^3.$$

This is larger by a factor of order A/a^2 than the momentum of any particle of the normal fluid, which will be of the order \hbar/a at most. Thus no collision with atoms of the normal fluid will cause such transitions to occur.

7. THE CRITICAL VELOCITY.

In this section we shall give an extremely tentative indication of the way in which a critical velocity of flow may arise. We shall assume that it is possible for the superfluid in a given tube to be partly in motion and partly at rest, with a spatial boundary between the moving part and the stationary part. We shall assume moreover that the wave functions at this boundary behave in the same way as at the boundary of the fluid (*i. e.* as in fig. 2), and that a surface energy σ exists there. On dimensional grounds we should expect

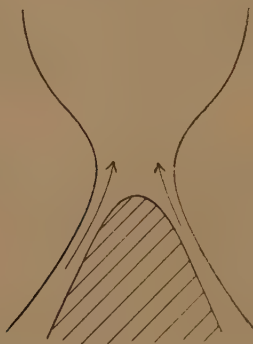
$$\sigma \sim \hbar^2/ma^4.$$

If a nucleus of stationary fluid exists, it will obviously be possible for it to expand, since atoms will continually be exchanged by means of transitions to and from the normal fluid, from the moving to the stationary part, with gain in energy. One will thus obtain persistent flow only in the absence of stationary nuclei. The energy gained when an atom is transferred from a part of the fluid moving with velocity v to the stationary fluid is $\frac{1}{2}mv^2$; thus any nucleus of stationary fluid will behave as if it had a pressure $\frac{1}{2}mv^2/a^3$ tending to make it expand.

Consider then helium II flowing in a capillary in which there is a constriction of diameter d , and let the velocity be v at that point. Then from the end of the capillary where it dips into stationary fluid we may expect a stationary nucleus to form, and spread up the tube. The way this might happen is indicated in fig. 6; the shaded area represents the stationary fluid. If the fluid in the tube is to stop completely, then the stationary nucleus must force its way up through the constriction. The surface tension will oppose this with a force of the order σ/d per unit area, so that the nucleus will spread and the fluid eventually come to rest if

$$\sigma/d < \frac{1}{2}mv^2/a^3.$$

Fig. 6.



This gives a critical velocity

$$v \sim \sqrt{(a^3 \sigma / md)}.$$

If we set here $\sigma \sim \hbar^2 / ma^4$, this gives

$$v \sim \hbar / m \sqrt{(ad)}.$$

This differs from the previous empirical formula \hbar/md , but for thin capillaries gives results of a similar order of magnitude.

Note added November 21st.—Certain predictions can be made from the theory presented in this paper. One is, of course, that the light isotope of helium, He^3 , will not show superfluidity, because the atoms obey the Fermi-Dirac statistics.

Another is as follows*. Imagine the tube (or concentric cylinders), shown in fig. 4 filled with helium I and rotated, and, while still in rotation, cooled below the λ -point. If the tube is then stopped the distribution of atoms is as in fig. 5 (a). If the tube is cooled still further, normal fluid atoms will condense into the state k_0 . Thus the angular momentum of the helium will increase. Since angular momentum is conserved, the tube if freely pivoted will start to rotate in the opposite direction.

* This test of the theory was suggested by Dr. Pippard.

REFERENCES.

- ANDRONIKASHVILI, E., 1946, *J. Phys. U.S.S.R.*, **10**, 201.
BIJL, A., DE BOER, J., and MICHELS, A., 1941, *Physica*, **8**, 655.
DINGLE, R. B., 1948, *Proc. Phys. Soc.*, **61**, 9.
FRENKEL, J., 1940, *J. Phys. U.S.S.R.*, **2**, 365.
KISTERMACHER, J., 1947, *Physica*, **13**, 81.
LANDAU, L., 1941, *J. Phys. U.S.S.R.*, **5**, 71; 1944, *Ibid.*, **8**, 1; 1947, *Ibid.*, **11**, 91.
LONDON, F., 1938, *Phys. Rev.*, **54**, 947.
MOTT, N. F., and JONES, H., 1936, *Theory of the Properties of Metals and Alloys*.
SCHIFF, L. I., 1941, *Phys. Rev.*, **59**, 839.
TISZA, L., 1947, *Phys. Rev.*, **72**, 838.

V. The Coefficients of Expansion of some Solid Solutions in Aluminium.

By W. HUME-ROTHERY, F.R.S., and T. H. BOULTBEE*,

The Inorganic Chemistry Laboratory, The University Museum, Oxford.

[Received June 24, 1948.]

I.

THE way in which the coefficient of expansion of a metal is affected by the solution of a second metal is not yet understood. E. A. Owen and E. W. Roberts (1939) used X-ray methods to measure the coefficients of expansion of solid solutions of cadmium, indium, tin and antimony in silver, and found that in all cases the solid solution had a greater coefficient of expansion than the pure metal. In these alloys the formation of the solid solution is accompanied by an expansion of the lattice, and for a given atomic percentage of solute the lattice expansion increases with the valency of the solute, whilst the coefficient of expansion also increases. It was suggested by one of us that solid solutions might show increased or decreased coefficients of expansion according as the lattice distortion was respectively an expansion or contraction. One of the objects of the present investigation was to test this hypothesis for solid solutions in aluminium, and for this purpose mean coefficients of expansion have been determined for the ranges $-50^{\circ}\text{C. to }+25^{\circ}\text{C.}$, and $+25^{\circ}\text{C. to }+200^{\circ}\text{C.}$

The use of X-ray methods for the determination of coefficients of expansion has also been discussed by Hume-Rothery and Reynolds (1938) and by A. J. Wilson (1941), but doubt has existed as to the accuracy which can be attained. A further object of the present work was to examine the reproducibility of results obtained by carefully controlled X-ray methods.

II. EXPERIMENTAL DETAILS.

(a) *Preparation of Specimen.*

The alloys used in the present work were prepared from metals of the highest purity, the details of which have been given in the paper by Axon and Hume-Rothery (1948). The British Aluminium Co. Ltd., must

* Communicated by the Authors.

again be thanked for presenting the super-purity aluminium from which the alloys were made. For experiments at $-50^{\circ}\text{C}.$, and at room temperature, many of the specimens used by Axon and Hume-Rothery (1948) were available, and the exact methods of preparation will be found in the paper by these authors. For work at $+200^{\circ}\text{C}.$ the filings were enclosed in thin silica capillaries, and the methods were the same as those of Hume-Rothery and Reynolds (1938).

(b) *Lattice Spacing Measurements.*

For experiments at room temperature use was made of a standard 19 cm. Unicam Camera, the 18 cm. low-temperature camera (L.T.C.) of Hume-Rothery and Strawbridge (1947), and the 9 cm. high-temperature camera (H.T.C.) of Hume-Rothery and Reynolds (1938). The temperature was observed carefully during the exposure, and the mean value usually lay in the range $18^{\circ}\text{--}20^{\circ}\text{C}.$ The lattice spacing was then determined by the standard extrapolation methods, and was corrected to a standard temperature of $25^{\circ}\text{C}.$ The agreement between results obtained from different cameras is discussed in Section III. In the work at low temperatures, the temperature was kept constant to within $\pm 0.3^{\circ}\text{C}.$ The mean temperature usually lay between -47° and -50° , and the results were corrected to a standard temperature of $-50^{\circ}\text{C}.$ In the high temperature work, the results were corrected to a standard temperature of $+200^{\circ}\text{C}.$, and the temperature during exposure was kept constant to within $\pm 0.3^{\circ}\text{C}.$ The details of the high temperature (1938) and low temperature (1947) techniques have already been described.

The experiments referred to above gave accurate spacings at -50° , $+25^{\circ}$, and $+200^{\circ}\text{C}.$, and from these values the mean coefficients of expansion $\alpha_{t_1}^{t_2}$ were determined, α being defined by the relation

$$\alpha_{t_1}^{t_2} = \frac{a_{t_2} - a_{t_1}}{a_{t_1}(t_2 - t_1)} \quad \dots \dots \dots (1)$$

Here a_{t_2} , and a_{t_1} , are the lattice spacings at temperatures t_2 and t_1 , where t_2 is the higher * temperature.

III. REPRODUCIBILITY OF RESULTS.

Two determinations of the lattice spacing of pure aluminium using filings annealed in alumina tubes inside evacuated glass tubes were made by one of the present authors (T.H.B.) and gave values (at $25^{\circ}\text{C}.$) of 4.04142 and 4.04141 kX units respectively. Two determinations made by the other author (W. H.-R.) gave values 4.04141 and 4.04146 kX

* At temperatures below $0^{\circ}\text{C}.$ some authors define α as being equal to $\frac{a_{t_2} - a_{t_1}}{a_{t_2}(t_2 - t_1)}$ where t_2 is the higher temperature. Since the expression (1) is almost always used for high temperature measurements, it seems more consistent to retain it for the low temperature range.

units. The experiments were all made in the 18 cm. L.T.C., and the mean value * 4.04142 kX is in fairly good agreement with the value $4.0413(4) \text{ kX}$ given by Axon and Hume-Rothery (1948) as the mean of nine experiments in the 19 cm. Unicam camera; these values lay between the extremes $4.0412(8)$ and $4.0414(2) \text{ kX}$.

In the course of the present work determinations were made of the lattice spacings of filings of pure aluminium which had been annealed in glass tubes without the precaution of an inner tube of alumina. The resulting lattice spacings were about 0.0002 kX units less than those of the above values, and this suggested a very slight contamination by silicon. In tracing this source of error, a number of duplicate experiments was made by one of us (T. H. B.) in which the conditions were kept as uniform as possible. Using the 18 cm. L.T.C. six independent experiments gave a mean value of 4.04123 kX , the total spread being 0.00009 kX , whilst three experiments in the 19 cm. Unicam camera gave a mean value of 4.04120 kX and a spread of 0.00004 kX . The above results suggest that the 19 cm. Unicam camera gives lattice spacings which are very slightly lower than those from the 18 cm. L.T.C., the difference being of the same order as the accuracy of the calibration of the cameras. All results in the present paper are referred to the L.T.C., and in places where it has been necessary to use results from the 19 cm. Unicam camera we have increased these by 0.00003 kX in order to make data from the two cameras as comparable as is possible.

The results for experiments at -50°C . are given in Tables I.-IV. and it will be seen that where more than one experiment was made the agreement is of the same order as for the data at 25°C . The same tables contain the data for $+200^\circ \text{C}$. together with experiments made to ensure that the specimens had not changed in composition during the heating. It will be seen that the H.T.C. values are in very good agreement with those from the other two cameras, but in view of the smaller diameter of the H.T.C. we do not think that an accuracy of more than $\pm 0.00005 \text{ kX}$ should be claimed.

IV. DATA FOR ALUMINIUM.

The results † for pure aluminium are given in Table I. and indicate that over the range -50°C . to $+200^\circ \text{C}$. the mean coefficient of expansion is 22.3×10^{-6} . This may be compared with the value 22.6×10^{-6} calculated from the equation given in the 'Metals Handbook' (1939) for bulk metal, and with the value 22.3×10^{-6} obtained by combining the results of Ebert (1928) in the range of 0°C . to -50°C . with the usually accepted

* We have preferred the value 4.04142 rather than 4.04143 because the results of T. H. B. were obtained at a later date and are probably more reliable than those of W. H. R.

† The present data at -50°C . are to be regarded as replacing those of Hume-Rothery and Strawbridge (1947) owing to an improvement in thermocouple calibration.

value 23.1×10^{-6} for the range 0° to $+25^\circ$ C. It is therefore clear that no detectable difference exists between the expansion of the lattice and of the bulk metal over the range -50° to $+25^\circ$ C.

Owing to accidents it was unfortunately not possible to determine the lattice spacing at 200° C. of filings annealed in alumina tubes. One sample of filings annealed in a glass tube gave a lattice spacing at 25° of 4.04137 kX, and values of 4.05900 kX at 200° C., and a repeat value of 4.04138 kX at 25° C. All these experiments were made in the H.T.C., and it is clear that any contamination with silicon must have been extremely small. The resulting mean coefficient of expansion α_{25}^{200} is 24.9×10^{-6} , and may be compared with the values 24.6×10^{-6} and 24.7×10^{-6} given in the 'Metals Handbook' for the range 20° to 200° C. for wrought aluminium of 99.996 per cent and cast aluminium 99.95 per cent purity respectively, both in bulk form*. It is again clear that the lattice and the bulk metal have the same coefficient of expansion to within the limits of the experimental methods.

TABLE I.
Data for Pure Aluminium.

Experiment No.	Camera	Temperature	Lattice Spacing	Remarks.
A7	L.T.C.	-60° C.	4.03468 kX	
14	L.T.C.	-50° C.	4.03469	
3	L.T.C.	25° C.	4.04141	See p. 72
2	L.T.C.	25° C.	4.04146	"
B47	L.T.C.	25° C.	4.04142	"
B48	L.T.C.	25° C.	4.04141	"

$$\alpha_{-50}^{+25} = 22.3 \times 10^{-6}$$

V. DATA FOR ALUMINIUM ALLOYS.

The data for the solid solutions of magnesium in aluminium are given in Table II., and in fig. 1 the values of α_{-50}^{+25} are shown as a function of the composition. The coefficient of expansion of aluminium is increased by the solution of magnesium, but the effect is not in direct proportion to the concentration. The first additions of magnesium cause a relatively smaller increase in the coefficient of expansion than is produced by further addition, and it is noteworthy that the lattice spacing composition curve of Axon and Hume-Rothery shows a curvature of the same kind. For the alloy containing 10.56 atomic per cent magnesium, the value of α_{25}^{200} is 26.7×10^{-6} and is thus considerably greater than that for aluminium. This alloy lies outside the limits of the α solid solution and it was feared that

* The X-ray results of A. J. Wilson (1942) for filings annealed at 600° give a value of $\alpha_{25}^{200} = 24.3 \times 10^{-6}$. Correspondence with Dr. Wilson shows that these filings were annealed in glass tubes.

precipitation of the β phase would occur on heating to 200° C. As will be seen from Table II. there was no detectable difference in the lattice spacings before and after treatment at 200° C., and it appears that precipitation does not occur in the short time of an X-ray experiment. For the

TABLE II.

Experiment No.	Alloy	Camera	Temperature	Lattice Spacing	Remarks
B35	2.36 at % Mg.	L.T.C.	-50° C.	4.04276 kX	
B69	2.36 "	L.T.C.	-50° C.	4.04275	
B70	2.36 "	L.T.C.	-50° C.	4.04275	
B34	2.36 "	L.T.C.	+25° C.	4.04965	
0	2.36 "	19 cm. Unicam	+25° C.	4.04960	*
$\alpha_{-50}^{+25} = 22.7 \times 10^{-6}$					
A2	6.36 at % Mg.	L.T.C.	-50° C.	4.06143	
B75	6.36 "	L.T.C.	-50° C.	4.06143	
B73	6.36 "	L.T.C.	+25° C.	4.06851	
B53	6.36 "	H.T.C.	+25° C.	4.06854	Value after experiment at 200° C.
B52	6.36 "	H.T.C.	+200° C.	4.08577	
$\alpha_{-50}^{+25} = 23.2 \times 10^{-6}$					
$\alpha_{25}^{200} = 24.7 \times 10^{-6}$					
B72	10.56 at % Mg.	L.T.C.	-50° C.	4.07823	
B77	10.56 "	L.T.C.	+25° C.	4.08613	
0	10.56 "	Unicam	+25° C.	4.08611	*
B55	10.56 "	H.T.C.	+25° C.	4.08617	Before heating to 200° C.
B58	10.56 "	H.T.C.	+25° C.	4.08616	After heating to 200° C.
B56	10.56 "	H.T.C.	+200° C.	4.10383	
B57	10.56 "	H.T.C.	+200° C.	4.10368	
$\alpha_{-50}^{+25} = 25.9 \times 10^{-6}$					
$\alpha_{25}^{200} = 26.7 \times 10^{-6}$					

6.36 atomic per cent magnesium alloy the value of α is 24.7×10^{-6} and is thus very slightly less than that for aluminium, although the difference is within the experimental error. There is thus increasing evidence that the first small additions of magnesium to aluminium produce an effect different from that of later additions.

The data for the solid solutions of zinc in aluminium are given in Table III., and the values of α_{-50}^{+25} are plotted in fig. 1. The coefficient of expansion of aluminium is increased by the solution of zinc, and the increase is very nearly if not exactly proportional to the atomic percentage of solute*.

The alloy containing 1.62 atomic per cent copper was examined in considerable detail. The data are given in Table IV., from which it will be seen that the coefficient of expansion α_{-50}^{+25} is very nearly the same

TABLE III.

Experiment No.	Alloy	Camera	Temperature	Lattice Spacing	Remarks
X	1.70 at% Zn	L.T.C.	-50° C.	4.03343	
8	1.70 „	L.T.C.	+25° C.	4.04036	
0	1.70 „	Unicam	+25° C.	4.04041	*
$\alpha_{-50}^{+25} = 22.9 \times 10^{-6}$					
X	3.63 „	L.T.C.	-50° C.	4.03164	
B76	3.63 „	L.T.C.	+25° C.	4.03881	
$\alpha_{-50}^{+25} = 23.7 \times 10^{-6}$					
61	5.9 „	L.T.C.	-50° C.	4.02942	
0	5.9 „	Unicam	+25° C.	4.03699	*
$\alpha_{-50}^{+25} = 25.1 \times 10^{-6}$					

TABLE IV.

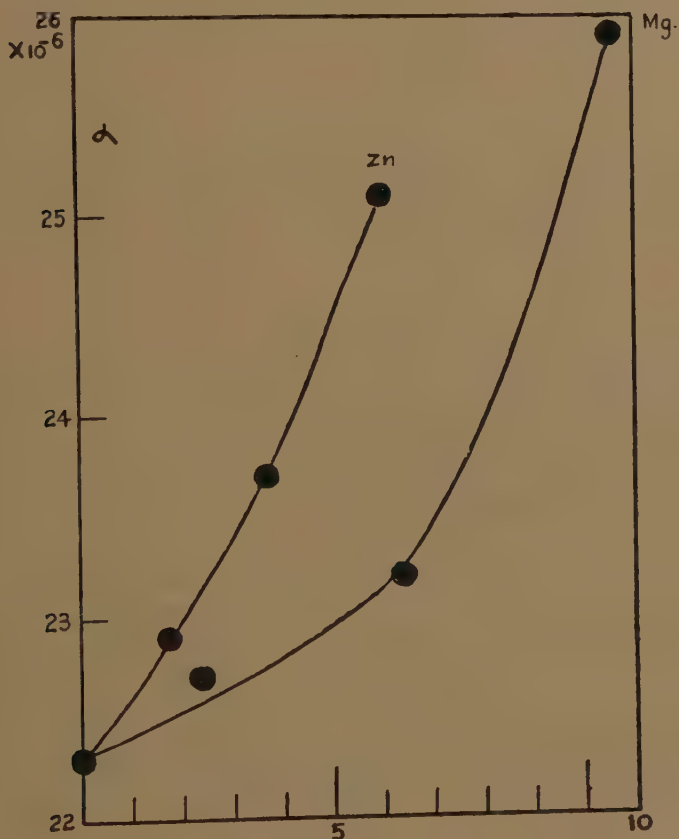
Experiment No.	Alloy	Camera	Temperature	Lattice Spacing in kX units	Remarks
B67	1.62 at % Cu	L.T.C.	-50° C.	4.02667	
B68	1.62 „	L.T.C.	-50° C.	4.02670	
X2	1.62 „	L.T.C.	+25° C.	4.03353	
0	1.62 „	Unicam	+25° C.	4.03354	*
B62	1.62 „	H.T.C.	+25° C.	4.03347	After heating to 200° C.
B60	1.62 „	H.T.C.	+200° C.	4.05110	See above
B61	1.62 „	H.T.C.	+200° C.	4.05104	
$\alpha_{-50}^{+25} = 22.5 \times 10^{-6}$					
$\alpha_{25}^{200} = 24.8 \times 10^{-6}$					

as that of aluminium. In Table IV. we include also a value of α_{25}^{200} which is again about the same size as that for aluminium. The behaviour of this alloy is puzzling because the composition 1.62 atomic per cent copper is outside the limits of the solid solution at 200° C. The filings had been quenched from the homogenous region before they were used for the room temperature and low temperature work, and it

* The curve in the figure could be replaced by a straight line without exceeding the experimental error.

would be expected that precipitation of CuAl_2 would occur on heating to 200°C . Actually, as can be seen from Table IV., there is no indication of any change in lattice spacing in room temperature photographs taken before and after the heating to 200°C ., from which it appears that precipitation did not occur. In a recent paper Elwood and Silcock (1948) have shown that at 548°C . the coefficient of expansion of aluminium is appreciably lowered by the solution of copper. In fig. 2 of the paper

Fig. 1.



In this Figure the mean coefficients of expansion between -50°C . and $+25^\circ \text{C}$. are plotted as functions of the composition for solid solutions of magnesium and zinc in aluminium.

of these authors, lattice spacing composition curves are given for different alloys over the range 300° to 600°C ., and examination shows that the difference in the slopes of these curves becomes less as the temperature falls, and that at 350° – 450° they are almost indistinguishable, and would indicate almost the same coefficient of expansion for the alloys and for the pure metal. The conclusions of the present work are therefore not in contradiction to those of Elwood and Silcock.

From Table V. it will be seen that the coefficient of expansion of aluminium is increased by the solution of 4.83 per cent of lithium, whilst 0.92 atomic per cent of silicon produces no change outside the limits of experimental error.

TABLE V.

X4	0.92 at % Si	L.T.C.	-50° C.	4.03288	
0	0.92 „	Unicam	+25° C.	4.03967	*
$\alpha_{-50}^{+25} = 22.4 \times 10^{-6}$					
X5	4.83 at % Li.	L.T.C.	-50° C.	4.03196	
0	4.83 „	Unicam	+25° C.	4.03930	*
$\alpha_{-50}^{+25} = 24.3 \times 10^{-6}$					

Note.—In Tables I. to IV. the values marked * are those of Axon and Hume-Rothery increased by 0.00003 to allow for the difference between the L.T.C. and Unicam Cameras (see pp. 75, 76).

DISCUSSION.

The results described above show clearly that the coefficient of expansion of aluminium is increased both by the solution of elements which produce an expansion (magnesium) and a contraction (zinc, lithium) of the lattice of the solvent. The increase in the coefficient of expansion is thus not necessarily associated with an expansion of the lattice, but appears to be a result of lattice distortion without reference to the sign of the distortion.

Comparison of the results given in the tables will show that the use of 18 cm. or 19 cm. cameras under carefully controlled conditions usually enables results to be obtained which are consistent to within 0.00005 kX units. The change in lattice spacing of pure aluminium between -50° and +25° C. is approximately 0.007 kX units. The determination of the coefficient of expansion involves the difference between two lattice spacings. It is very improbable that a single determination of a lattice spacing will differ by more than 0.00005 kX from the "true apparent value" * in the cameras concerned, so that even where single determinations are made coefficients of expansion can be compared over this temperature range to an accuracy of the order 1 per cent, whilst greater accuracy should be possible if repeat experiments are made and a mean value is taken. This accuracy compares favourably with that obtained by methods involving use of bulk metal. With the 9 cm. H.T.C. the accuracy is about one half as great for the same temperature range. In the present work the H.T.C. range was from +25° C. to +200° C.,

* We use the expression "true apparent value" to denote the result from a given camera on the assumption that the thermocouple is calibrated correctly.

and the mean coefficients of expansion can therefore be compared to a degree of accuracy at least as great as that for the range -50°C. to $+25^{\circ}\text{C.}$ in the L.T.C.

The absolute accuracy of the measurements depends on the calibrations of the camera and thermocouple. Of these the former will affect the individual lattice spacings, but will have a negligible effect on the coefficient of expansion which involves the difference between two lattice spacings. In the calibration of the thermocouple an accuracy of $\pm 1.0^{\circ}\text{C.}$ is fairly easily obtained, but an accuracy of $\pm 0.1^{\circ}\text{C.}$ is much more difficult, and at present this is the weak point in the method. It is, however, doubtful whether the errors involved are greater than those in most methods using a bulk specimen, since in these there is the further difficulty of keeping the whole length of the specimen at the same temperature. The metal aluminium is favourable for accurate measurements of the expansion coefficient by X-ray methods because, with copper $K\alpha$ radiation, it gives a strong (511, 333) reflection at approximately 82° backed by a strong (422) reflection at 69° , so that extrapolation methods can be used with certainty. So far as the extrapolation methods are concerned, the error in lattice spacing measurements may be expected to increase roughly in proportion to the extent of the extrapolation, and hence in proportion to the value of $\cos^2 \theta$ for the line of highest angle. For a metal whose line of highest angle is 76° this source of error will be roughly three times that for aluminium. If, however, all experiments are carried out with the same specimen in exactly the same adjustment *, the accuracy with which the coefficient of expansion can be determined will not fall off so rapidly. The X-ray method has the advantage that only a small amount of material is needed, and difficulties of obtaining exact uniformity of composition over a bulk specimen are avoided. It seems therefore, that for homogeneous alloys of cubic structure the X-ray method of determining coefficients of expansion is to be recommended in cases where the X-ray diffraction pattern gives rise to a line of reasonably high angle. It will of course be understood that the method gives the coefficient of expansion of the lattice, and takes no account of possible effects of flaws, cavities, or other irregularities in the crystal structure which may affect the expansion of the bulk metal.

ACKNOWLEDGMENTS.

The authors must express their thanks to Professor C. N. Hinshelwood, F.R.S., for laboratory accommodation and many other facilities which have greatly encouraged the present research. Grateful acknowledgment is also made to the Council of the Royal Society, and to the British Non-Ferrous Metals Research Association for financial assistance towards the expenses of their research work.

* It is necessary to emphasize this point because if a specimen becomes slightly out of centre between two experiments the slopes of the extrapolation curves will not be the same.

REFERENCES

- OWEN, E. A., and ROBERTS, E. W., 1939, *Phil. Mag.* (7), **39**, 294.
 HUME-ROTHERY, W., and REYNOLDS, P. W., 1938, *Proc. Roy. Soc. A*, **167**, 25.
 WILSON, A. J., 1941, *Proc. Phys. Soc.*, **53**, 235; 1942, **54**, 487.
 AXON, H. J., and HUME-ROTHERY, W., 1948, *Proc. Roy. Soc. A*, **193**, 1.
 HUME-ROTHERY, W., and STRAWBRIDGE, D. J., 1947, *J. Sci. Instruments*, **24**, 89.
 EBERT, H., 1928, *Zeit. Physik*, **47**, 712.
 ELLWOOD E. C., and SILCOCK, J. M., 1948, *J. Inst. Metals* (May).
Metals Handbook, 1939, 1241. (This is a standard American book.)

VI. *An Application of the Theory of Quasi-Properties to the Treatment of Anomalous Strain-Stress Relations.*

By G. W. SCOTT BLAIR, M.A., D.Sc., F.R.I.C., F.Inst.P. and
 J. E. CAFFYN, B.Sc., A.Inst.P. *

The National Institute for Research in Dairying, University of Reading †.

[Received February 9, 1948.]

Summary.

The idea of quasi-properties proposed in an earlier paper is extended to cover anomalous strain-stress relations.

Just as the Newtonian definition of equality of time intervals may be said to depend on the constancy of the magnitude of the vector c , so any definition of equality of forces, if independent of Newtonian time, must depend on the directional character of c . In the light of these implicit definitions, it is seen that the significance to be attached to fractional differential coefficients of strain with respect to stress is somewhat different from that put forward in the earlier paper for fractional strain-time differentials.

Certain experiments on relaxation of plastics at constant stress were selected on account of their exceptional divergence from the Nutting relaxation law, and it was shown that for the less serious anomalies, an equation integrated from the strain-stress fractional differential in some cases fitted the data better than the corresponding strain-time equation. For extreme anomalies, an integrated form of the previously proposed double fractional differential equation, reduced to four variables, is shown to fit the data admirably.

It is claimed that, when the aim is to relate the complex rheological behaviour of materials to the results of tests in which their firmness is

* Now Lecturer in Physics, The Durham Colleges, University of Durham.

† Communicated by Dr. S. Whitehead.

assessed subjectively by handling, it is better to express the data in terms of quasi-properties rather than by the classical methods used to study the structure of materials.

The circumstances under which it is practicable to express the Nutting equation and its fractional differential derivatives in a simple dimensional form are discussed.

Introduction.

In an earlier paper (G. W. Scott Blair, B. C. Veinoglou and J. E. Caffyn 1947) it was suggested that the behaviour of complex materials under stress " might be described in terms of entities which are not strictly ' physical properties.' These so-called ' quasi-properties ' range from entities hardly distinguishable from dimensionally true physical properties to concepts which are much less clearly defined." It was pointed out that the constancy of velocity of bodies moving independently of the action of other bodies * or of light in free space depends on the definition of the equality of time intervals inherent in Newtonian physics.

In the rheology of many complex systems the entities which are found to be constant can be expressed either in terms of fractional differential coefficients of strain (σ) with respect to stress (S) and time (t), or by using a non-Newtonian definition of equality of time intervals. In the latter case, fractional differentials could be eliminated at least for time differentiation. It was pointed out that, while this explains the appearance of fractional differential coefficients when Newtonian time is used, there is everything to be said for keeping the Newtonian definition in practice even at the expense of the more complicated mathematical treatment involved. It was therefore not proposed to use a non-Newtonian time scale.

It would seem that these ideas have given rise to misgivings among some of those who have read our work ; though, for lack of published comment and since few precise criticisms have been made directly to the authors, it has proved difficult to find out exactly where the difficulties lay.

So far as we can judge, misgivings about the treatment we have proposed may be broadly classified under three headings, (1) the fact that our treatment does not lead to any understanding of the structure of the materials or of their molecular configurations, (2) our use of entities whose dimensions depend on the nature of the material, (3) our use of fractional differentials and especially our description of a fractional differential coefficient as " intermediate " between the zero and first differential coefficients. We will deal with these points in turn.

(1) In this matter, one is reminded of the controversy between Boltzmann and Mach. The treatment used must clearly depend on the purpose

* For the sake of brevity we shall refer to this simply as " the velocity of light " or " c " without repeating the alternative reference to a moving body.

of the investigation. We wish here to state categorically that it has never been our intention to elucidate the molecular (solid) structure of materials, but to describe the phenomenological relations between stresses, strains, and times in materials which are generally so complex that any straining to which they are subjected for experimental purposes immediately changes their rheological properties. This is the case with most thixotropic systems, and we must regard the materials studied by C. F. Goodeve and G. W. Whitfield (1938) as exceptional in that they apparently maintain a steady equilibrium stress for each constant rate of shear. It is a first principle of physics that physical properties cannot be measured by methods which change them appreciably during and as a result of the measurement. In such cases we are obliged to measure processes and not properties or to forego all measurements on such systems. As stated in our earlier paper we are dealing with non-equilibrium conditions.

In spite of the phenomenological nature of our approach, our results are far from being purely empirical or unrelated to fundamental concepts, but the fundamental concepts are not molecular but are concerned with the judgment of the rheological behaviour of materials by handling. Such judgments are said to be subjective in the sense that they relate to states of feeling but the particular class with which we are concerned are reproducible as statistical distributions and may thus be defined and assessed quantitatively.

(2) We find it hard to take the dimensional argument seriously*. It is, of course, true that in the Nutting equation (P. G. Nutting 1921)

$$\psi = S^{\beta} \sigma^{-1} t^k, \quad (1)$$

ψ is not, in the usual sense of the term, a 'physical property.' It has dimensions which depend on the nature of the material, so that two materials cannot be directly compared from the magnitudes of ψ alone, the relative magnitudes of the ψ -values depending on the values of S and t chosen for comparison. The equation is not, as has sometimes been alleged, dimensionally inhomogeneous, since ψ has the dimensions $[ML^{-1}T^{-2}]^{\beta}[T]^k$, but this does not affect the issue†. There can be no reason to object to the use of the power-law equation on dimensional grounds unless the mistake were made of claiming an unjustifiable status for ψ , which indeed we have never done.

Not only can the equation be used as an empirical explanation of the facts, but we have been able to show (G. W. Scott Blair and F. M. V. Coppen 1943) that the exponent k is a psychophysical entity which can be evaluated from the subjective handling tests to give a value agreeing

* Professor H. Dingle deals with this question in another paper in this issue (p. 94).

† Compare H. Dingle's illustration of dimensional inhomogeneity:—"48 inches + 3 hours = 4 feet + 180 minutes" in which, of course, the exponents of M, L and T differ on the two sides of the equation (Phil. Mag., xxxiii. 321, 1942).

within the limits of experimental error with that obtained by the values calculated in a certain manner from the results of purely physical experiment.

In fact ψ is now never used in practice to specify the behaviour of materials. It is better to take as a measure of intensity of firmness, the volume, $1/\zeta$ of the $S : \sigma : t$ figure integrated over the effective ranges of S and t (Scott Blair 1943). This volume is given by

$$\frac{1}{\zeta} = \int_0^{S_{\max}} \int_0^{t_{\max}} \sigma \, dS \cdot dt$$

which, on substituting from the Nutting equation, gives

$$\log \zeta = \log (\beta + 1) + \log (k + 1) - \log S_{\max} - \log t_{\max} - \log \sigma_{\max}$$

when σ_{\max} is the maximum strain corresponding to $S_{\max} : t_{\max}$, or

$$S_{\max} t_{\max} \cdot \zeta = (\beta + 1)(k + 1) / \sigma_{\max}, \quad \dots \dots (1a)$$

a dimensionless equation in which ζ has the dimensions of a fluidity ($M^{-1}LT$). The parameter ζ is not, of course, a fluidity nor is it a "physical property" in the dimensional sense of the term. This means that for materials for which it is found experimentally that $\log \sigma / \log t$ and $\log \sigma / \log S$ curves are linear, if we select any suitable values of S_{\max} and t_{\max} , and measure experimentally the corresponding strain σ_{\max} , and if we also measure the (constant) slopes of the $\log \sigma / \log t$ and $\log \sigma / \log S$ curves,

we can calculate a ratio $\frac{(\beta + 1)(k + 1)}{\zeta} = S_{\max} t_{\max} \sigma_{\max}$. From the values of

β , k and ζ we can express the whole of the $S : \sigma : t$ relations of the material, up to the values of S_{\max} and t_{\max} chosen, for the whole range over which the Nutting equation is found empirically to hold under the conditions of test. If we consider a parallelepiped having sides of length S_{\max} , t_{\max}

and σ_{\max} , $\frac{1}{\beta + 1} \cdot \frac{1}{k + 1}$ is the fraction of this parallelepiped contained by the $S : \sigma : t$ figure and describes the nature of the material, reducing, for example, to $\frac{1}{4}$ for an elastic solid and to $\frac{1}{2}$ for a viscous fluid. The ratio $S_{\max} \cdot t_{\max} / \sigma_{\max}$ represents the intensity of firmness, *e.g.* elastic modulus and viscosity for prototype systems. Thus if we are given ψ_A and ψ_B for two materials A and B, we know next to nothing about their comparative behaviour, but if we are also given k_A , k_B , β_A and β_B , we can compare their rheological behaviour over the whole range for which the equation holds.

Such entities as $\boxed{\psi : k : \beta}$ taken together are (a) useful in comparing materials, (b) reproducible by different observers in different laboratories, (c) in some cases give the same magnitudes when derived from entirely different processes, but since it is a process which is being described, this cannot be counted on in all cases. (Even in the measuring of simple physical quantities, alternative processes of measurement often do not give precisely the same result as Dingle (1942) has stressed.)

It has been shown experimentally (Scott Blair and Coppen, *loc. cit.*) that these entities which we call "Quasi-properties" are quantitatively linked with the entities (Gestalten) by which people judge the consistency of materials in handling them; judgments which are (whether we like it or not) of very great importance in industry.

(3) The third criticism concerns the sense in which it is true to say that a fractional differential coefficient, such as $d^{\frac{1}{2}}\sigma/dt^{\frac{1}{2}}$, is "intermediate" between $d\sigma/dt$ and $\sigma = d^0\sigma/dt^0$.

As was shown in the earlier paper (Scott Blair, Veinoglou and Caffyn, *loc. cit.*), and as will be further shown below, the fractional differential equations are empirically justified by the really excellent fit of the experimental data to the series equations which they give on integration.

We regard the fractional differentials themselves as constituting an ordered sequence of mathematical operations such that, in the simple case quoted above, three operations of the type $d^{\frac{1}{2}}\sigma/dt^{\frac{1}{2}}$ would be equivalent to a single unit differentiation. This is a matter of definition and it is clear that it is the operation which constitutes the "intermediacy" and not the algebraic or numerical expressions derived from it. It is well known, of course, that the *magnitudes* derived by the fractional differentiation sometimes lie outside the boundaries set by the zero and first differentiations (R. Harper 1947). It is not these magnitudes which are intermediate.

It is very much hoped that these brief remarks will allay any remaining doubts about the soundness of the methods which we have proposed.

Complex Strain-Stress Relations.

In the earlier paper, the definition of equality of forces was not considered, nor was any adequate explanation given for the use of fractional differential coefficients of strain with respect to stress. It was suggested that the general entity intermediate between length and velocity, having dimensions in T, $>0 < 1$ should be called "motion."

It was pointed out by Scott Blair, Veinoglou and Caffyn (*loc. cit.*), that the Newtonian definition of equality of time units is a definition of a constant velocity—*i. e.* of the constancy of the magnitude of the vector c (Nutting 1921). The definition of equality of forces, however, does not define a constant first differential of length with respect to force or of strain with respect to stress, *i. e.* Hookean behaviour. If we seek a definition of equality of forces which does not itself depend on acceleration, *i. e.* on the Newtonian time scale, we shall find that, though such a definition may be expressed in more than one way, it must ultimately rest on the concept of the straight line or of motion without change of direction, *i. e.* on the directional part of the vector c . For example, one might define two forces A and B as equal when, acting in the same direction on a body, a third force C can be found which would balance either of them, in the sense that the body will not change its motion as a result of the simultaneous actions of either A—C or B—C. It follows that if the body is of finite size, A (or B) and C must be acting in a straight

line—i. e. following the path of the vector c . If the body becomes a point, there must be no change of direction at this point, the forces acting there “in opposite directions.” It would not seem to have been noticed previously that the Newtonian definitions of time and force equalities may be described as ultimately depending on the constancy of magnitude and direction respectively of the vector c . (See also Dingle 1939.)

If an analogous treatment is applied to rheological quantities, considered independently and not supposed to arise from the application of mechanical laws to models of solid structure, then there is an essential difference between $d\sigma/dt$ and $d\sigma/dS$. In a Newtonian fluid there is, in fact, a constant relative velocity between sheared elements since $d\sigma/dt$ is a constant with which a time scale for the variation of the rheological properties could be associated. Unlike some other rheologists, we are prepared to use the term strain to include fluid flow of a rod as well as plastic and elastic deformations, the strain being defined as $\sigma = \int_{l_0}^t \frac{dl}{l}$, where l_0 is the initial length of the test piece and l , the length after time t . The rate of extension is then given by $d\sigma/dt$ from which the rate of shear can be calculated. For constant stress this implies for a Newtonian fluid, a constant velocity relative to the immediate surroundings so that $d\sigma/dt$ has the same status as a velocity. The quantity $d\sigma/dS$ cannot be associated with a scale of force because equality of two forces in rheology does not depend on the constancy of a rheological motion.

Hookean bodies are Hookean by reason of the approximate cancellation of curvatures of their attractive and repulsive force-curves for small deformations and there is no reason why, for large deformations and complex materials, the constant factor should be a whole number differential coefficient of strain with respect to stress rather than some fractional coefficient. In the simplest non-Hookean case, described by C. Bach (1888), the stress is proportional to a power of the strain. This fact does not tell us anything about the molecular structure of materials, but we would repeat that this is not our present interest.

Moreover, so long as relaxation experiments follow a rapid extension or compression, our relaxation curves generally follow a simple power law quite closely. Following slow extension, it is natural that it takes longer for the Nutting quasi-equilibrium for relaxation, described in the earlier paper, to be established.

It is the purpose of the present paper to apply integrated forms of the fractional differential equation relating σ to S and the general double fractional differential equation relating σ to S and t , to data obtained by measuring stress relaxation in complex plastics at constant strain following a relatively slow straining at constant stress. But before proceeding to this, we must return to the question put in the earlier paper as to whether $\sigma : S$ anomalies, in which time does not appear to enter, could be accounted for in terms of a non-Newtonian time scale. It is now clear that, in some cases at least, they could be so accounted for.

Although the conditions are entirely different, it is worth recalling that, according to the General Theory of Relativity, an acceleration which could be interpreted by an outside observer in terms of a difference in time scales could equally justifiably be recorded by a second observer, attached to the "accelerating" system, as a static reading on a spring balance. Since this is the case in one sphere of physics, there is no reason why it should not be attempted in quite a different connection.

This method of accounting for stress anomalies in terms of time scales was, in fact, carried out with some success in the earlier paper, but it will be recalled that there were a few rare cases of relaxation experiments in which, in order to obtain an adequate fit, the calculations became extremely laborious.

There are, of course, a number of alternative ways in which such anomalies might be treated—*e. g.* the definition of strains in a non-Euclidean space—but such devices, though perfectly legitimate in themselves, would not seem to further our immediate purpose.

Alternatively, it would be feasible to suggest that each material has a characteristic time scale for each stress, but this would be difficult to interpret as an explanation of Bachian behaviour where time does not appear as a variable.

It is better to regard $\sigma : S$ anomalies in the same manner as we have previously regarded $\sigma : t$ anomalies, only with the difference already pointed out.

Use and Integration of the Fractional Differential Equations.

In our earlier paper, three principal equations were proposed in addition to the Nutting equation already quoted:—

$$\psi = S^\beta \sigma^{-1} t^k, \quad (1) \quad (\text{The Nutting Equation—loc. cit.})$$

$$\chi_1 = S \div \left| \frac{\partial^\mu \sigma}{\partial t^\mu} \right|, \quad (2) \quad \left. \begin{array}{l} \text{Partial differential Equations. (The} \\ \text{symbol "F" indicates fractional} \\ \text{differentiation.)} \end{array} \right\}$$

$$\chi_2 = t \cdot \left| \frac{\partial^e S}{\partial c^e} \right|, \quad (3)$$

$$X = \left| \frac{d^e S}{d \left(\left(\frac{d^\mu \sigma}{dt^\mu} \right)^e \right)} \right|, \quad (4) \quad \text{The General Equation.}$$

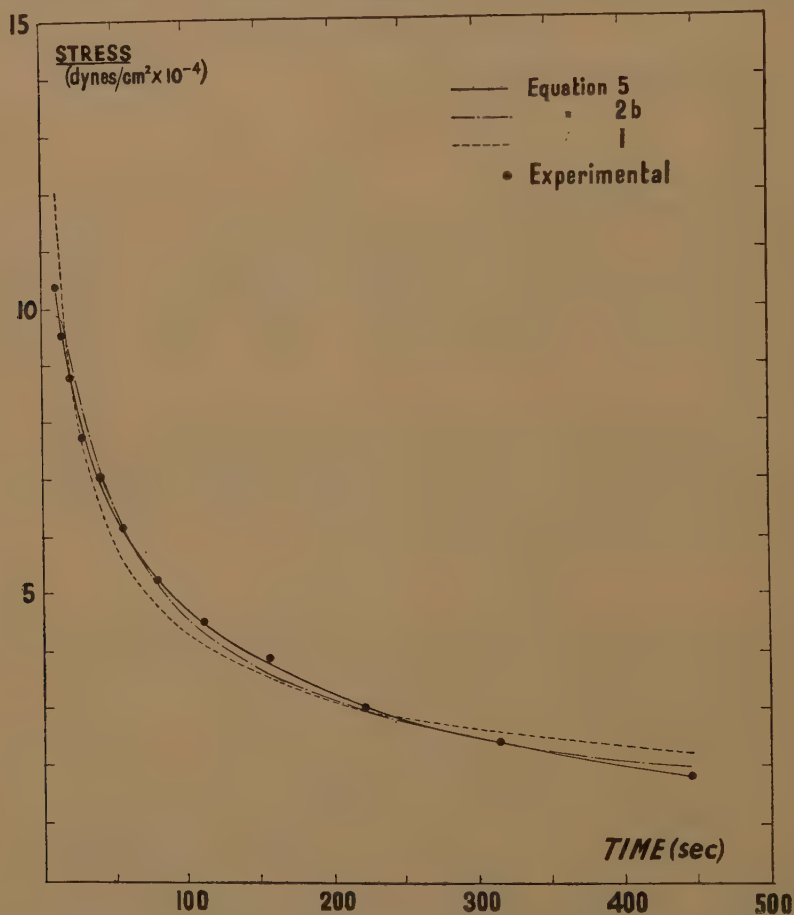
We now propose to re-write equations (2), (3) and (4) somewhat differently. Since we are primarily concerned with relaxation, it would be better to re-define χ_1 and χ_2 in terms of S^β and t^k so that equation (2) is written

$$\chi_1 = S^\beta \div \left| \frac{\partial^\mu \sigma}{\partial t^\mu} \right|. \quad \dots \dots \dots (2a)$$

The observables of rheological experiments are the deformations from which strains are derived and the changes in deformation occurring in time or associated with different stress conditions. Moreover, in any expression relating $\sigma : S$ the exponents of S and t are of the same sign

high temperatures (35°–120° C.) used by M. D. Stern and A. V. Tobolsky (1946) than at the temperatures of 15° C. and 18° C. used by us.

Tables I *a* and I *b* show the deviations of the experimental points from the best fitting curves calculated from equations (1), (2 *b*) and (5) respectively for samples of two different materials. Table II shows the deviations of the experimental points from the best fitting curves calculated from equations (1), (2 *b*) and (3 *b*) respectively for a different material. Table III shows the deviations of the experimental points from the best fitting curve calculated from equation (5) for another material which showed the most anomalous relaxation curves from amongst our data.



These materials are the plastic (unspecified 2), butyl 2, cellulose acetate + phthalate, and plastic (unspecified 5) of Table II of our earlier paper (Scott Blair, Veinoglou and Caffyn—*loc. cit.*).

Typical relaxation curves calculated from equations (1), (2 *b*) and (5), together with the experimental points for a single experiment, are shown in the figure.

We are indebted to Dr. L. J. Comrie of the Scientific Computing Service for calculating most of the data given in the Tables.

TABLE Ia.
Plastic (unspecified 2).

log t	Sample I				Sample II				Sample III			
	log S_{exp}	d_1	d_{2b}	d_5	log S_{exp}	d_1	d_{2b}	d_5	log S_{exp}	d_1	d_{2b}	d_5
1.00	4.900	-34	-8	+2	5.205	-44	+12	+5	5.335	-47	+25	+7
1.18	4.880	-10	+7	-2	5.180	-15	-10	-4	5.300	-24	-18	-7
1.30	4.860	+2	-10	-3	5.160	+4	-12	-5	5.280	-2	-23	-5
1.48	4.835	+21	0	+3	5.125	+23	-7	0	5.235	+12	-26	-8
1.65	4.790	+20	-2	-1	5.080	+32	+2	+3	5.195	+31	-8	0
1.78	4.765	+26	+8	+4	5.040	+30	-6	+3	5.165	+42	+11	+8
1.95	4.710	+15	+7	+1	4.980	+25	-12	+5	5.110	+46	+30	+13
2.08	4.670	+7	+6	0	4.930	+13	+12	+5	5.070	+48	+47	+20
2.26	4.610	-9	+4	-2	4.850	-13	+5	+4	4.990	+26	+49	+18
2.38	4.550	-38	-14	+1	4.770	-55	-22	-4	4.790	-68	-90	-37

S_{exp} = Experimental Stress (Dynes/cm.²).
 d_1 = Deviation of S_1 from S_{exp} etc.

TABLE I b.

Butyl 2.

log t	Sample I				Sample II				Sample III			
	log S_{exp}	d_1	d_{2b}	d_5	log S_{exp}	d_1	d_{2b}	d_5	log S_{exp}	d_1	d_{2b}	d_5
1.00	5.018	-64	+21	+3	5.350	-56	+19	+4	5.511	-61	+18	+2
1.15	4.980	-34	-12	-4	5.312	-31	-11	-3	5.484	-32	-7	-2
1.30	4.945	-3	-16	+2	5.272	-7	-20	-3	5.448	-7	-18	-4
1.45	4.890	+9	-25	-6	5.230	+15	-16	-1	5.409	+15	-16	-3
1.60	4.850	+36	-6	+6	5.182	+30	-7	+3	5.369	+33	-4	+3
1.75	4.790	+43	+3	+3	5.124	+36	0	+1	5.316	+40	+3	+2
1.90	4.720	+39	+6	-1	5.065	+41	+11	+3	5.256	+40	+9	+1
2.05	4.655	+41	+19	0	5.000	+40	+20	+2	5.196	+40	+20	+1
2.20	4.590	+43	+36	+8	4.923	+26	+20	0	5.123	+25	+19	-1
2.35	4.484	+4	-13	-7	4.841	+8	+16	-2	5.039	+1	+10	-6
2.50	4.384	-30	-2	-3	4.740	-29	-4	-3	4.956	-23	+3	-1
2.65	4.260	-86	-41	+5	4.630	-76	-33	+3	4.844	-76	-32	+1

 S_{exp} = Experimental Stress (Dynes/cm²). d_1 = Deviation of S_1 from S_{exp} etc.

TABLE II.

Cellulose Acetate + Phthalate.

log t	Sample I			Sample II			Sample III			Sample IV			Sample V		
	log S_{exp}	d_1	d_{2b} d_{3b}	log S_{exp}	d_1	d_{2b} d_{3b}	log S_{exp}	d_1	d_{2b} d_{3b}	log S_{exp}	d_1	d_{2b} d_{3b}	log S_{exp}	d_1	d_{2b} d_{3b}
1.00	6.959	-26	+6 -1	6.984	-42	+1 -7	6.878	-11	+8 +4	6.912	-11	+5 +1	7.005	-11	+6 +4
1.18	6.928	-8	0 +3	6.959	-16	-4 0	6.832	-8	-3 0	6.863	-9	-4 -3	6.960	-3	+1 +4
1.30	6.900	-4	-6 +3	6.936	-7	-9 +2	6.800	-8	-7 -2	6.835	-3	-2 +1	6.922	-6	-6 -1
1.48	6.855	-1	-10 -2	6.904	+9	-3 +5	6.764	+5	0 +3	6.789	+2	-1 +1	6.870	-6	-9 -5
1.60	6.827	+3	-8 -1	6.876	+14	-2 +4	6.720	-6	-12 -7	6.754	+1	-3 -1	6.836	-5	-10 -6
1.70	6.800	+3	-8 -4	6.852	+16	-1 +3	6.702	+3	-2 -2	6.729	+4	0 0	6.813	+1	-3 -3
1.78	6.783	+8	-3 -2	6.827	+13	-3 -1	6.687	+9	+4 +2	6.707	+4	+5 0	6.793	+5	0 -1
1.88	6.762	+13	+4 0	6.806	+18	+5 +1	6.661	+10	+6 +1	6.682	+8	+5 +1	6.765	+6	+2 -2
1.95	6.742	+12	+4 0	6.787	+18	+6 +1	6.640	+8	+5 +1	6.661	+6	+4 +2	6.745	+6	+3 0
2.02	6.724	+13	+6 0	6.765	+15	+6 -1	6.622	+9	+6 +1	6.640	+5	+4 0	6.727	+9	+6 +2
2.08	6.708	+13	+8 0	6.747	+13	+6 -2	6.604	+8	+25 0	6.625	+7	+7 +2	6.714	+13	+12 +5
2.26	6.662	+15	+16 +3	6.687	+1	+2 -6	6.552	+4	+6 -1	6.570	+3	+5 0	6.657	+9	+11 +2
2.38	6.620	+6	+11 +1	6.646	-8	0 -4	6.512	-3	0 -2	6.524	-9	-5 -5	6.615	+2	+5 +1
2.56	6.552	-14	-1 0	6.584	-22	-3 0	6.462	-4	+3 +2	6.473	-9	-2 -1	6.552	-9	-2 0
2.68	6.492	-42	-24 -1	6.542	-32	-6 +8	6.410	-24	-13 -1	6.437	-11	-2 +4	6.500	-26	-16 -3

 S_{exp} = Experimental Stress (Dynes/cm.²); d_1 = Deviation of S_1 from S_{exp} etc.

TABLE III.
Plastic (unspecified 5).

Sample I			Sample II			Sample III		
$\log t$	$\log S_{\text{exp}}$	d_7	$\log t$	$\log S_{\text{exp}}$	d_7	$\log t$	$\log S_{\text{exp}}$	d_7
1.15	4.880	0	1.00	5.125	+35	1.26	5.145	+2
1.28	4.810	+3	1.18	5.045	-21	1.45	5.065	-1
1.38	4.730	-9	1.30	4.990	-28	1.58	5.000	-5
1.46	4.680	+1	1.40	4.940	-27	1.76	4.915	+2
1.53	4.630	+6	1.54	4.860	-15	1.83	4.875	0
1.59	4.570	-2	1.65	4.790	-6	1.89	4.845	+3
1.64	4.522	-3	1.74	4.735	+8	1.94	4.815	+3
1.69	4.490	+7	1.81	4.680	+14	1.99	4.790	+5
1.73	4.435	-4	1.88	4.625	+13	2.20	4.655	+3
1.77	4.410	+7	1.93	4.590	+25	2.44	4.440	-11
1.81	4.350	-5	1.98	4.540	+21	2.60	4.305	+7
1.84	4.325	-2	2.02	4.485	+9			
1.87	4.295	+1	2.08	4.450	+28			
1.90	4.270	+7	2.13	4.400	+29			
1.92	4.225	-2	2.22	4.280	+6			
2.00	4.140	+5	2.35	4.070	-29			
2.06	3.965	-31	2.45	3.895	-41			
2.16	3.830	+18	2.54	3.745	-29			
			2.61	3.445	+29			

S_{exp} = Experimental Stress (dynes/cm²).

d^2 = Deviation of S_7 from S_{exp} .

Discussion and Conclusions.

We conclude that to describe the behaviour of those comparatively rare materials which, in relaxation, are far from obeying the Nutting Law, the introduction of one extra term in either equation (2 *b*) (*i. e.* equation (8) in the earlier paper) or equation (3 *b*) gives a very much better and generally fully adequate fit to the experimental points; and that in the cases of most extreme anomaly, where equations (2 *b*) and (3 *b*) are inadequate, the general equation (5) fits admirably when modified as described above to give four parameters. There are cases for which the stress anomaly is evidently of greater significance than the time anomaly, and for which the use of equation (3 *b*) with its two terms gives a better fit than that of equation (2 *b*) with the same number of terms. We feel that this excellent agreement on a large body of experimental data fully justifies our use of these dimensionally complex equations for the purpose we have in view.

If we are prepared to consider only qualities as distinct from intensity, the dimensional difficulty does not arise. Indeed, in its differential form, the Nutting equation can readily be shown to be one of the simplest expressions of dimensionless ratios relating $\sigma : S : t$ for anomalous systems.

A test-piece has a characteristic length (l_0) before a test is done on it. It does not have a characteristic time or (normally) a characteristic initial stress. In order that deformations may be expressed in terms of an entity which, like t and S , is zero at the start of the experiment, strain is defined in its simplest form as $\Delta l/l_0$ or sometimes as $\Delta l/l$, where Δl is the increase in length during the time t . If, in order to get dimensionless ratios, we express increments of stress and time likewise as fractions of their total magnitudes, we have the expressions $\Delta S/S$ and $\Delta t/t$. But, since strain, being zero at the start of the experiment, is analogous to S and t rather than to $\Delta S/S$ and $\Delta t/t$, we must repeat the process on σ and write $\Delta\sigma/\sigma$.

For anomalous systems, there can be no simpler possible relation between these dimensionless ratios (reduced to infinitesimals) than

$$\frac{d\sigma}{\sigma} = \beta \frac{dS}{S} + k \frac{dt}{t}, \quad (1b)$$

which is the differentiated form of the Nutting equation.

The Nutting equation is, as we have shown, the first term of a double series equation and we are indebted to Mr. L. L. Whyte for pointing out to us (private communication) that the factors in subsequent terms of each of these series differ from one another dimensionally by a time and a stress respectively. Thus, for example, in equation (2 b), B/A represents a time characteristic of the material and all values of t might be expressed as ratios of this time.

We could write equation (5) as

$$\sigma = AA' t^{k'} S^{\beta'} + AB' t^{k'} S^{\beta'-1} + BA' t^{k'-1} S^{\beta'} + BB' t^{k'-1} S^{\beta'-1}. \quad . . . (5a)$$

Considering only the most important ratios we could write

$$\psi' = \frac{1}{\sigma} \left(\frac{t}{t_0} \right)^{k'} \left(\frac{S}{S_0} \right)^{\beta'} \left(1 + \frac{S_0}{S} + \frac{t_0}{t} \right), \quad (6)$$

where $S_0 = AB'/AA'$ and $t_0 = BA'/AA'$.

If so many complex systems did not show such excellent approximation to the simple Nutting law, equation (6), being dimensionless, might well prove most useful. In practice, however, t_0 and S_0 have to be calculated from the deviations from the Nutting equation and, in the great majority of our experiments, these deviations are far too small to give anything like reasonable accuracy.

For systems which obey the Nutting equation, to express stress and time in the dimensionless forms S/S_0 and t/t_0 would involve postulating special values of S and t , supposedly characteristic of the material, for whose existence there is no evidence. The evaluating of the dimensionless equation (1 a) was only made possible by selecting arbitrary limits for S and t , but these were limits of the field of interest and not characteristics of materials.

It remains true, therefore, that for most complex systems tested by us at room temperature either under constant stress or strain, behaviour

is best expressed for our purposes in terms of quasi-properties. The practical applications of the fractional differential equations to problems of assessing the firmness of such materials by handling, have recently been reviewed in a psychophysical study by Harper (*loc. cit.*), who has shown that it is possible to select values of the exponent for the fractional differentiation of strain with respect to time which can quantitatively account for the time anomalies in comparing the firmness of rubber and bitumen by handling, earlier reported by Scott Blair and Coppen (*loc. cit.*).

Acknowledgments.

Our best thanks are due to the British Electrical and Allied Industries Research Association for permission to publish this work and to many friends for helpful advice. We would especially mention Drs. M. Reiner, J. Topping, S. Whitehead, P. White and Professor H. Dingle, who have given us much help and encouragement. We also wish to thank Dr. B. C. Veinoglou, who carried out many of the original experiments from which the data discussed in this paper have been selected.

One of us (J. E. C.) is indebted to the Agricultural Research Council for a grant.

References.

- BACH, C., 1888, *Z. Ver. deutsch. Ingen.*, **32**, 192.
 DINGLE, H., 1939, *Nature, Lond.*, **144**, 888; 1942, *Phil. Mag.*, **33**, 321.
 GOODEVE, C. F., and WHITFIELD, G. W., 1938, *Trans. Faraday Soc.*, **34**, 511.
 GREEN, M. S., and TOBOLSKY, A. V., 1946, *J. Chem. Phys.*, **14**, 80.
 HARPER, R., 1947, *J. Psychol.*, **60**, 554.
 NUTTING, P. G., 1921, *Proc. Amer. Soc. Test. Mater.*, **21**, 1162.
 SCOTT BLAIR, G. W., 1943, *Nature, Lond.*, **152**, 412.
 SCOTT BLAIR, G. W., and COPPEN, F. M. V., 1943, *Amer. J. Psychol.*, **56**, 234.
 SCOTT BLAIR, G. W., VEINOGLOV, B. C., and CAFFYN, J. E., 1947, *Proc. Roy. Soc. A*, **186**, 69.
 STERN, M. D., and TOBOLSKY, A. V., 1946, *J. Chem. Phys.*, **14**, 93.

VII. *On the Dimensions of Physical Magnitudes* (Seventh Paper : *A Paradox in Dimensional Theory*).

By HERBERT DINGLE, A.R.C.S., D.I.C., D.Sc.*

[Received February 9, 1948.]

IN some recent papers †, G. W. Scott-Blair and his collaborators have given evidence of the existence of a relation in the field of rheological measurements which presents a new problem in dimensional theory.

* Communicated by the Author.

† For example, *Proc. Roy. Soc. A*, clxxxix. p. 69 (1947), where references to others are given.

The relation, in the form originally given by P. G. Nutting *, is

$$\psi = S^\beta \sigma^{-1} t^k, \quad \dots \quad (1)$$

where S is the stress and σ the strain of a material which is neither a Hookean solid nor a Newtonian fluid, after a time t from the beginning of the experiment when $\sigma=0$. The quantities ψ , β and k are constant for a given material over a considerable range of the variables, but vary from one material to another. In this respect they resemble elasticity and viscosity, and so would appear to have equal right to be considered characteristic "properties" of the material. They differ from elasticity and viscosity, however, and from all other generally recognized properties of materials, in that it appears to be impossible to assign to ψ at least definite dimensional formulæ of the ordinary type. For example, taking, as usual, M, L and T to be fundamental measurements, the dimensions of ψ come out as $[M^\beta L^{-\beta} T^{k-2\beta}]$ if the principle of dimensional homogeneity is to be satisfied. The dimensions thus vary from one substance to another, since β and k vary with the substance.

This creates a problem whatever view of dimensional theory one takes. On the older view, which in a series of recent papers † I have attempted to discredit, the dimensional symbol characterizes the *property* and not the substance having that property, so that the dimensions of density, for example, being $[ML^{-3}]$, every density, no matter how, where or when manifested, must necessarily have those dimensions. On the view which I proposed to substitute for this, the dimensional symbol characterizes a *process of measurement*, not any particular application of that process, so that if the measurements involved in the determination of a density are indicated by $[ML^{-3}]$, that symbol represents every measurement of density, no matter in what physical system it is made. In Scott-Blair's systems, however, the same physical quantity, ψ , or the same process of measurement, may have dimensions $[M^2 L^{-2} T]$ for one substance and $[ML^{-1} T^{\frac{1}{2}}]$ for another (these indices are imaginary, and do not necessarily correspond to any actual materials). Accordingly his work has been criticized by those who hold the older view of dimensions, on the ground that the quantities which he has found useful experimentally are not true properties of substances, and it has been maintained that he must find some other way of expressing his results which is not open to the dimensional objection. Partly on this account, Scott-Blair sometimes refers to his quantities as "quasi-properties." Naturally I am not much impressed by this objection, and I am perfectly content to leave to those who take the older view the task of explaining how something that is not a property of a substance can exhibit itself experimentally as though it were, but I cannot evade the obligation to consider the

* Journ. Franklin Inst. xcxi. p. 679 (1921); Proc. Amer. Soc. Test. Mater. xxi. p. 1162 (1921).

† Phil. Mag. xxxiii. p. 321 (1942); xxxiii. p. 692 (1942); xxxiv. p. 588 (1943); xxxv. p. 296 (1944); xxxv. p. 588 (1944); xxxvii. p. 64 (1946).

meaning of such quantities as ψ , β and k in terms of my own view of dimensions. On the one hand, they certainly represent measurements in the sense that the processes by which they are determined are definite and definable independently of the substances to which they refer. On the other hand, the dimensional symbol, supposed to characterize each definite and independent process uniquely, is indefinite and not unique. How can these two facts be reconciled with the operational view of dimensions?

Let me premise two remarks. First, in the earliest of my papers on this subject I wrote * :—"The fact that alternative processes of measurement of the same physical quantity do not inevitably give the same result has led thinkers of positivist schools to discard the idea of a physical quantity altogether and to grant meaning only to the process and result of measurement in each case While this view is doubtless formally correct, we shall not adopt it since it introduces unnecessary elaboration and, moreover, tends to obscure relationships which it is the object of physics to bring to light. We shall still speak of physical quantities measurable by different methods, but since we shall never introduce a physical quantity apart from the measurement of it and shall, further, regard each process of measurement as distinct from every other, we shall preserve the rigour of the true positivist view without suffering from its inconvenience." Quite apart from the present problem I have gradually come to regret this compromise since it has given rise to more misunderstanding than it is worth. Apparently one cannot touch pitch without becoming defiled, and some critics have found it impossible to speak of physical quantities (or physical properties) without giving them a naively realistic "existence" which is independent of the process by which we determine them. Accordingly, I would now categorically withdraw all significance from the expression "physical quantity" or "physical property," and regard every magnitude considered in physics simply as the numerical result of a specific process of measurement. The object of physical research is then to find relations between the results of different measurements. These relations are expressed by equations. Every algebraic symbol in such an equation stands for the result of a measurement, and the corresponding dimensional symbol in the related dimensional equation stands for the process, simple or complex, which defines that measurement. This indicates the point of view from which I propose to contemplate the problem aroused by Scott-Blair's work.

Secondly, I am not at all concerned here with any aspect of that work beyond the dimensional problem already stated. It is for experts in the field of rheology to determine whether his experimental work is accurate and his conclusions sound. It is sufficient for my purpose that they *could* be sound, *i. e.* the results of definite processes of measurement might stand to one another in the relation given by (1) or some

* Phil. Mag. xxxiii. p. 324 (1942).

similar relation. In order, therefore, to isolate the discussion in this paper from all questions of rheological practice and theory, and at the same time to reduce the essential problem to its simplest form, I propose to discuss a purely imaginary case in which that problem appears as uncomplicated as possible by accompanying details. Let us then consider one of the earliest of all successfully solved physical problems—the law of falling bodies. We know that this is given by $s = \frac{1}{2}gt^2$, but we will simply take it in the form

$$s = \lambda t^\alpha, \quad (2)$$

where λ and α are independent constants. It is only by repeated experiments on many bodies that we know that λ and α are independent of the body experimented on, and we must therefore admit the possibility, so far as the principles of physical investigation are concerned, that they might have varied with the body used, otherwise we must account Galileo and Newton inadmissibly deficient in intelligence for experimenting with different substances. It is therefore a possibility which should be covered by any general theory of measurement and by any acceptable theory of dimensions that, when s and t are measured, say, in centimetres and seconds respectively, λ and α might be 100 and 2 for gold, 200 and $3/2$ for silver, 300 and $5/4$ for lead, and so on. In such a case their values would be characteristic for particular materials, and our problem is to assign to each of them a dimensional symbol that shall definitely represent the process by which it is measured and remain the same for all applications of that process whatever may be the body on which it is performed.

We may take [L] and [T] as fundamental magnitudes for this purpose, so that there is no question of how to measure s and t ; we are concerned only with λ and α . To find their dimensions we have therefore to ask ourselves how we measure them. The answer is simple. Take two values of s and t during a single fall—namely, s_1, t_1 and s_2, t_2 —and solve the following equations for λ and α :—

$$\left. \begin{aligned} s_1 &= \lambda t_1^\alpha, \\ s_2 &= \lambda t_2^\alpha \end{aligned} \right\} (3)$$

The result is

$$\left. \begin{aligned} \alpha &= \frac{\log s_1/s_2}{\log t_1/t_2}, \\ \log \lambda &= \frac{\log s_1 \log t_2 - \log s_2 \log t_1}{\log t_2/t_1} \end{aligned} \right\} (4)$$

From this α is obviously dimensionless since both s_1/s_2 and t_1/t_2 are pure numbers, and the dimensions of λ are [antilog (log L log T)]. Hence we have obtained symbols which represent the processes of measurement and are independent of the particular application of those processes, and so have solved our problem.

In solving it, however, we have raised another. We have no experience of dimensional symbols involving transcendental functions like those obtained for λ ; indeed, it is generally understood that in order to satisfy the principle of dimensional homogeneity—which simply expresses the requirement that a physical relation, if true for one set of units of measurement of the magnitudes involved, shall be true for all sets of units—the dimensional expression for every term in a physical relation must take the form of a simple product of powers of the fundamental magnitudes adopted. The proof is given, for example, in Bridgman's book, 'Dimensional Analysis,' which represents essentially, though not necessarily in the form of expression, the view I am taking. It would appear, therefore, that by satisfying the requirement that the dimensional symbol shall represent the process of measurement, we have been forced to violate the principle that physical relations shall be true for all units of measurement.

Nevertheless, that is not so, as we may see by making a change of units. Suppose we write $s=aS$, $t=bT$, where S and T are the measures of length and time in the new units. (S , of course, must not be confused with S in (1) where it represents stress). Then α , being dimensionless, is unchanged, and we obtain from (4), by substitution,

$$\log \lambda = \frac{\log a S_1 \log b T_2 - \log a S_2 \log b T_1}{\log T_2/T_1},$$

which reduces to

$$\log \lambda = \frac{\log S_1 \log T_2 - \log S_2 \log T_1}{\log T_2/T_1} + \frac{\log a \log T_2/T_1 + \log b \log S_1/S_2}{\log T_2/T_1}. \quad (5)$$

But in order that our relation shall survive the change of units we must have

$$S = \Lambda T^\alpha, \quad (6)$$

where

$$\log \Lambda = \frac{\log S_1 \log T_2 - \log S_2 \log T_1}{\log T_2/T_1}. \quad (7)$$

The relation between λ and Λ must therefore be

$$\left. \begin{aligned} \log \lambda &= \log \Lambda + \frac{\log a \log T_2/T_1 + \log b \log S_1/S_2}{\log T_2/T_1} \\ &= \log \Lambda + \log a - \alpha \log b, \end{aligned} \right\} \quad (8)$$

$$i. e. \quad \lambda = \Lambda a b^{-\alpha}. \quad (9)$$

Hence the change which λ , defined as in (4), must undergo when the units are changed, in order that the relation (2) shall remain true, is the same as if λ had dimensions $[LT^{-\alpha}]$, for with these dimensions we should

obviously have (9) satisfied. We may, accordingly, write the dimensions of λ as either $[\text{antilog}(\log L \log T)]$ or $[LT^{-\alpha}]$, for these expressions are equivalent. The latter form is, of course, the obvious one, exemplified, for instance, by the ordinary dimensional form for g , viz. $[LT^{-2}]$.

The paradox is thus resolved. It is true that the dimensional symbol represents the process of measurement and is the same for all applications of the process; and it is also true that the dimensional symbol satisfies the condition that physical relations shall be independent of the units of measurement, which requires them to be expressible in terms of simple products of powers of the fundamental magnitudes. What is not true is that the form of the symbol which meets the first requirement is necessarily the same as that which meets the second. There was never any reason to suppose that it was, for Bridgman's proof of the necessity for the simpler form does not demand that the indices occurring in that form shall be uniquely determined for each physical relation. It has, however, I think, always been tacitly assumed that this must be so, and the type of relation exemplified by the Nutting equation therefore represents an enlargement of the scope of the hitherto contemplated physical relations. This seems to me to have some importance in the theory of measurement.

So far as Scott-Blair's work is concerned, it follows that no objection lies against it on dimensional grounds. The results he obtains must be admitted as legitimately expressed possibilities; their actual validity, as I have said, is not here under consideration.

VIII. *On the Theory of Ferromagnetic Resonance.*

By D. POLDER, H. H. Wills Physical Laboratory, University of Bristol
(Now at Philips Research Laboratories, Eindhoven) *.

[Received August 26, 1948.]

Summary.

Kittel's classical theory of resonance in a magnetized ferromagnetic material is generalized; a phenomenological theory is obtained in which the dependence on the shape and the dimensions of the specimen no longer occurs. As an application the propagation of a plane electromagnetic (micro-) wave in a magnetized ferromagnetic medium is discussed. A quantum theory of the resonance phenomenon is given on the basis of a generalized Bloch spin-wave theory. An ambiguity in the use of the term gyromagnetic ratio is pointed out. It is suggested that the anomalous g -value found by resonance experiments may be explained by the combined effect of an electric crystal field and spin-orbit interaction.

* Communicated by Prof. N. F. Mott, F.R.S.

§ 1. Introduction.

WHEN a ferromagnetic material is magnetized up to its saturation value M_0 by means of a constant external field and a small additional magnetic field of a high frequency perpendicular to the constant field is applied, it has been observed that the induced alternating magnetic moment shows interesting resonance phenomena at a frequency ω_0 , roughly equal to the precession frequency of a free electron spin in the external field. In a field of 3000 Gauss, for instance, ω is of the order of magnitude of 10^{10} cycles/sec. Kittel (1947) has been able to give an explanation of the effect with the aid of a classical model, previously used by Landau and Lifshitz (1935) in a different connection. In its simplest form the model is the following:

1. The atoms of a ferromagnetic substance are magnetic tops; the magnetic moment and the angular momentum of one top are parallel vectors with constant absolute values of β and J . The gyromagnetic ratio is $\gamma = \beta/J$.

2. The forces exerted on the tops are: the force exerted by the external magnetic field, the magnetic dipole-dipole forces between the magnetic tops and forces between nearest neighbours representing the exchange interaction, which tends to align the magnetic moments of all tops.

3. The equation of motion for a top is: $d\mathbf{J}/dt = \mathbf{T}$, where \mathbf{T} is the torque exerted on the top.

Although this model is much too simple to account for many important features of ferromagnetism, it proves to be quite adequate for the description of ferromagnetic resonance. Let us consider a small ellipsoidal specimen of a non-conducting ferromagnetic material in which the atoms are arranged in a cubic lattice. The principal axes of the ellipsoid are the x , y and z axes. The constant external field H_0 and the magnetization M_0 are in the direction of the z -axis. When a small alternating external field $h_x^0 \exp(i\omega t)$ is applied, it is correct to assume that at a given time all tops have the same small deviations from their equilibrium positions, provided that the dimensions of the ellipsoid are very small in comparison with the wavelength of the high-frequency field. The assumption is consistent with our model, since in the case of a very small ellipsoid, the external fields can be regarded as homogeneous fields and the dipole-dipole interactions can be represented by a homogeneous field $H_i = (4\pi/3 - N_i)M_i$ ($i = x, y, z$) acting on the magnetic dipoles of the tops. Here N_x, N_y and N_z are the three demagnetization factors; \mathbf{M} is the magnetization. The exchange forces can be completely ignored, since the alignment of the tops is maintained during the motion. For the part \mathbf{m} of the magnetization that is proportional to h_x^0 one finds (Kittel 1948):

$$\left. \begin{aligned} & [\gamma^2 \{H_0 + (N_x - N_z)M_0\} \{H_0 + (N_y - N_z)M_0\} - \omega^2] m_x \\ & \qquad \qquad \qquad = \gamma^2 M_0 \{H_0 + (N_y - N_z)M_0\} h_x^0, \\ & [\gamma^2 \{H_0 + (N_x - N_z)M_0\} \{H_0 + (N_y - N_z)M_0\} - \omega^2] m_y = i\omega \gamma M_0 h_x^0. \end{aligned} \right\} \quad (1)$$

Several remarks can be made in connection with these equations. First of all we observe that the resonance frequency depends on the shape of the body. This is essentially due to the long-range character of the dipole-dipole forces between the magnetic tops. For a sphere the resonance frequency is equal to γH_0 , but for a flat disc with its short axis in the y -direction it is equal to $\gamma(B_0 H_0)^{\frac{1}{2}}$, where $B_0 = H_0 + 4\pi M_0$. Secondly, in the model we have ignored the possibility of the existence of crystal-anisotropy forces, which favour the orientation of the magnetic moments in the direction of special crystallographic axes. It is not difficult to account for these forces by introducing formally an internal magnetic field in the model. It is found then that the anisotropy forces affect the resonance frequency. In this paper we shall disregard the effect and assume that the forces are negligible; the point has been discussed by Kittel (1948) and Snoek (1947). Thirdly, the formulae cannot possibly describe precisely the experimental results, since in the fundamental equation of motion no terms occur which account for energy dissipation in the ferromagnetic material and we know from experiments that dissipation takes place. It is possible to introduce formally damping terms in the equation of motion (Kittel 1947) but the choice of the terms is somewhat ambiguous and, moreover, it seems doubtful (*cf.* §5) if a simple description of the shape of the resonance curve can be given in this way. Therefore we shall provisionally omit a discussion of the damping mechanism and its possible effect on the resonance frequency.

In this paper we shall first give an extension of the classical theory to cases where the motions of the different tops are not necessarily identical. We are thus led to a phenomenological description of the behaviour of a saturated ferromagnetic material in the microwave region (§2), which is in principle applicable to bodies of any size and shape. In §3 we shall apply the phenomenological theory to two special problems: the propagation of microwaves in a ferromagnetic medium and the behaviour of a small non-conducting ellipsoid in an alternating magnetic field. In §4 we shall make an attempt to verify the use of the classical equations of motion for the description of ferromagnetic resonance by means of a quantum mechanical calculation of the scattering of electromagnetic waves from a small ferromagnetic ellipsoid.

Several authors (Yager and Bozorth 1947, Griffiths 1946) have published the results of their experiments on ferromagnetic resonance. The results of the experiments of Yager and Bozorth (1947), which have been carried out on a conducting ferromagnetic material (supermalloy), are the most suitable for comparison with the theory. In supermalloy crystal-anisotropy forces are negligible and the sharpness of the experimental resonance curve indicates that the damping mentioned above is very small. When it is assumed that the classical theory is also applicable to conducting materials (*cf.* §2) and when the geometry of Yager and Bozorth's experimental arrangement is taken into account one may expect the resonance frequency ω_0 to be equal to $\gamma(B_0 H_0)^{\frac{1}{2}}$. Here γ should

be substantially equal to the gyromagnetic ratio of the electron spins as determined by gyromagnetic experiments. Gyromagnetic experiments usually give a value for γ a few per cent smaller than $2\mu_B/\hbar$ (μ_B is a Bohr magneton) while the experiments of Yager and Bozorth lead to a value about 10 per cent larger than the free spin value $2\mu_B/\hbar$. In §5 it is suggested that the discrepancy is due to spin-orbit interaction.

§2. *Phenomenological Theory.*

The equations (1) could easily be deduced from the classical model only by virtue of the assumption that all tops perform the same motion. This assumption was justified since we had chosen a small body of a favourable shape. Let us now consider the classical model of a ferromagnetic body of arbitrary shape and size. We shall assume that the atoms are arranged in a cubic lattice. In the presence of a constant field only, there will exist one configuration of the axes of the tops, for which the net torque exerted on each individual top is zero. We shall call the direction of the axis of a top in this situation the equilibrium direction of the top. The equilibrium directions of the tops are not necessarily identical. We shall assume that the high frequency fields cause small periodic deviations of the axes of the tops from their equilibrium positions and that at a given time the deviations of different tops are not necessarily identical. We shall further assume that the differences between the equilibrium directions and the differences between the deviations from these directions for any two neighbouring tops are so small that the exchange forces can be completely neglected. This condition will be formulated more precisely at the end of this section. On these assumptions the magnetic field acting on one special top is given by $\mathbf{H} + 4\pi\mathbf{M}/3$. Here \mathbf{H} and \mathbf{M} are the macroscopic field quantities as they occur in Maxwell's equations and they are to be measured in the part of the material where the top we are considering is present. The equation of motion of the top then is :

$$\gamma^{-1} \frac{d\boldsymbol{\beta}}{dt} = \boldsymbol{\beta} \times (\mathbf{H} + 4\pi\mathbf{M}/3). \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The equation is consistent with the condition that the absolute value of $\boldsymbol{\beta}$ is a constant. Since $\mathbf{M} = N_0\boldsymbol{\beta}$ (N_0 is the number of atoms per cm.³) we can write :

$$\frac{d\mathbf{M}}{dt} = \gamma\mathbf{M} \times \mathbf{H}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Let us suppose that in the equilibrium situation the magnetization and the magnetic field in the part of the material we are considering are given by the parallel vectors \mathbf{M}_0 and \mathbf{H}_c . Then $\mathbf{h} = \mathbf{H} - \mathbf{H}_c$ and $\mathbf{m} = \mathbf{M} - \mathbf{M}_0$ are small quantities periodic in time. We introduce an orthogonal system of axes ξ, η, ζ with the ζ -axis parallel to \mathbf{M}_0 . When we consider only

constant terms and terms linear in small quantities in the equation of motion we have

$$\left. \begin{aligned} i\omega m_{\xi} &= \gamma(m_{\eta} H_c - M_0 h_{\eta}), \\ i\omega m_{\eta} &= \gamma(M_0 h_{\xi} - m_{\xi} H_c), \\ m_{\zeta} &= 0, \end{aligned} \right\}, \quad \dots \quad (4)$$

or

$$\left. \begin{aligned} m_{\xi} &= \frac{\gamma^2 M_0 H_c}{\gamma^2 H_c^2 - \omega^2} h_{\xi} - \frac{i\omega \gamma M_0}{\gamma^2 H_c^2 - \omega^2} h_{\eta}, \\ m_{\eta} &= \frac{\gamma^2 M_0 H_c}{\gamma^2 H_c^2 - \omega^2} h_{\eta} + \frac{i\omega \gamma M_0}{\gamma^2 H_c^2 - \omega^2} h_{\xi}, \\ m_{\zeta} &= 0. \end{aligned} \right\} \quad \dots \quad (5)$$

Therefore we can write for $\mathbf{b} = \mathbf{h} + 4\pi \mathbf{m}$:

$$\left. \begin{aligned} b_{\xi} &= \mu h_{\xi} - i\alpha h_{\eta}, \\ b_{\eta} &= i\alpha h_{\xi} + \mu h_{\eta}, \\ b_{\zeta} &= h_{\zeta}, \end{aligned} \right\} \quad \dots \quad (6)$$

with

$$\mu = \frac{\gamma^2 H_c B_c - \omega^2}{\gamma^2 H_c^2 - \omega^2}, \quad \alpha = \frac{4\pi M_0 \gamma \omega}{\gamma^2 H_c^2 - \omega^2} \quad \dots \quad (7)$$

With equations (6) and (7) we have obtained the relation between the periodic parts of \mathbf{B} and \mathbf{H} in a saturated ferromagnetic medium, assuming the applicability of the classical model. (It should be noticed that the tensor which relates \mathbf{b} and \mathbf{h} may be a function of x, y and z .) The formulation of the classical theory with the aid of (6) and (7) is convenient since all problems in which periodic phenomena are concerned can now be treated with the usual phenomenological method, *i. e.* by using Maxwell's equations. A simple calculation (*cf.* §3) shows, for instance, that Kittel's expressions for the induced magnetization in a small ellipsoid (equation (1)) can be obtained in this way.

However, the equation (6), without the additional equations (7), have a more general significance than our derivation with the aid of the classical model might suggest. In an isotropic material, completely magnetized in the ζ -direction, the linear relation between \mathbf{b} and \mathbf{h} that is the most general linear relation from a point of view of symmetry considerations, is exactly given by equations (6). For the only condition to be satisfied by the permeability tensor which relates \mathbf{b} and \mathbf{h} is the rotational symmetry around the ζ -axis. As long as we do not refer to a special model α and μ may be arbitrary complex quantities. In the limit $\omega \rightarrow 0$, however, the tensor should reduce to a symmetrical tensor, *i. e.* $\alpha \rightarrow 0$ when $\omega \rightarrow 0$. Further, from considerations on conservation of energy it is easy to deduce

that μ and α are real quantities in a case where there is no energy dissipation in the medium. The special values for α and μ valid for the classical model (equations (7)) are in agreement with these statements.

Therefore the equations (6) provide a simple and quite general basis for the phenomenological description of periodic electromagnetic phenomena in a saturated isotropic ferromagnetic material. Strictly speaking we should remember that a ferromagnetic substance is never isotropic, but has cubic symmetry at the most. This means that the permeability tensor will have a more complicated form when the magnetization M_0 is arbitrarily directed with respect to the crystallographic axes. We believe, however, and it is suggested by the classical calculations given above, that at least in cubic crystals in which crystal-anisotropy forces are negligible, the equations (6) are good approximations of the actual phenomenological relations between \mathbf{b} and \mathbf{h} . It may be pointed out that our phenomenological description, just as all theories which involve the use of macroscopic field quantities \mathbf{D} , \mathbf{E} , \mathbf{B} and \mathbf{H} , breaks down as soon as appreciable changes of the field quantities occur over distances which are of the order of the interatomic distance a .

The simplest way to give a phenomenological description of periodic phenomena in dielectric, possibly conducting, ferromagnetic materials is undoubtedly to use equations (6) for the relation between b and h and to introduce a complex isotropic dielectric constant ϵ in Maxwell's equations for the relation between \mathbf{D} and \mathbf{E} . From the point of view of symmetry considerations, however, much more complicated relations between the field quantities are possible, *e. g.* an anisotropic ϵ dependent on the direction of M_0 . For the sake of simplicity we shall not consider these possibilities in this paper. With regard to the behaviour of dielectric or conducting ferromagnetic material, it should be pointed out that the special values of μ and α given by equations (7) were derived from a classical model of a non-dielectric, non-conducting substance. Therefore it is questionable whether equations (7) are still useful to predict resonance frequencies in dielectric and conducting materials. In view of the absence of a detailed theory of ferromagnetic resonance in these substances, we shall suppose that the equations (7) are applicable, but it must be emphasized that a further hypothesis is introduced by doing so.

Finally, we have to discuss more precisely the condition that the exchange forces should be negligible in the classical model. The exchange forces can be represented by a term $-\sum_{k,l} U_{kl} \mathbf{J}_k \mathbf{J}_l / J^2$ (summation over nearest neighbours) in the potential energy between the tops. Here U_{kl} is of the order of magnitude of $k\Theta$, Θ being the Curie-temperature. First of all we require that in the equilibrium configuration the direction of \mathbf{J} of any top practically coincides with the direction (ζ) of the field \mathbf{H} acting on that top, *i. e.* $J_{\xi}^{(eq)}$ and $J_{\eta}^{(eq)}$ must be small quantities. When we consider $J_{\xi}^{(eq)}$ and $J_{\eta}^{(eq)}$ for different tops as slowly varying functions of the spacial coordinates of the tops, the torque due to the exchange forces exerted on one top is of the order of $U\alpha^2/J$ times the second derivatives $J''_{\xi,\eta}$ of $J_{\xi}^{(eq)}$ and

$\mathbf{J}_{\eta}^{(eq)}$ with respect to x , y and z . In equilibrium this torque is equal to the torque exerted by the field \mathbf{H} . Since the latter torque is of the order $\beta H \mathbf{J}_{\xi, \eta}^{(eq)}/J$, we can write the condition as :

$$1 \gg J_{\xi, \eta}^{(eq)}/J \approx U a^2 J_{\xi, \eta}' / \beta H J.$$

When we take $H \approx 3000$ Gauss, $\Theta = 1,000^\circ \text{ K.}$, $a^2 = 10^{-15}$, $\beta = 10^{-20}$ we have

$$J_{\xi, \eta}' / J \ll 10^{11}.$$

Secondly, we require that in the case of a periodic motion of the tops, the periodic torque exerted on a top by the exchange forces should be small compared with the periodic torque exerted by the magnetic field \mathbf{H} . The first torque is of the order $U a^2 \dot{\mathbf{j}}_{\xi, \eta}' / J$ (where $\mathbf{j} = \mathbf{J} - \mathbf{J}^{(eq)}$) and the second is $\dot{\mathbf{j}}_{\xi, \eta} \beta H / J$. Therefore the exchange forces are negligible if

$$\dot{\mathbf{j}}_{\xi, \eta}' \cdot \dot{\mathbf{j}}_{\xi, \eta} \ll \beta H / U a^2 \approx 10^{11}.$$

Roughly speaking the condition means that the wavelength of the periodic motion of the tops in the material should be much larger than 10^{-5} cm.

§3. Applications.

The first application of the phenomenological theory we want to discuss is the propagation of plane electromagnetic waves in an infinitely large ferromagnetic medium, which is homogeneously magnetized in the z -direction. We seek a solution of Maxwell's equations in which the field quantities \mathbf{b} , \mathbf{h} , \mathbf{E} and \mathbf{D} are proportional to $\exp(i\omega t - 2\pi i(\mathbf{s}\mathbf{r})/\lambda)$, where \mathbf{s} is a unit vector indicating the direction of propagation of the wave. In this case Maxwell's equations reduce to

$$n\mathbf{E} \times \mathbf{s} = -\mathbf{b}, \quad n\mathbf{h} \times \mathbf{s} = \mathbf{D}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where $n = 2\pi c/\omega\lambda$. When we assume $\mathbf{D} = \epsilon\mathbf{E}$ we have

$$\epsilon\mathbf{b} = n^2\{\mathbf{h} - \mathbf{s}(\mathbf{h}\mathbf{s})\}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

With the aid of equations (8) and (9) we easily deduce that for any given value of the angle ϑ between \mathbf{s} and the z -axis there are two values for n^2 which satisfy (6) and (9). They are given by

$$n_{\pm}^2 = \frac{\epsilon(\mu^2 - \mu - \alpha^2) \sin^2 \vartheta + 2\mu \pm \{(\mu^2 - \mu - \alpha^2)^2 \sin^4 \vartheta + 4\alpha^2 \cos^2 \vartheta\}^{\frac{1}{2}}}{(\mu - 1) \sin^2 \vartheta + 1}. \quad (10)$$

The two corresponding solutions of Maxwell's equations represent two ellipsoidally polarized waves propagating in the same direction, but with different velocity. In the case $\vartheta = \pi/2$ we have two linearly polarized waves with

$$n_-^2 = \epsilon, \quad n_+^2 = \epsilon(\mu^2 - \alpha^2)/\mu. \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

In the case $\vartheta = 0$, we have two circularly polarized waves with

$$n_-^2 = \epsilon(\mu - \alpha), \quad n_+^2 = \epsilon(\mu + \alpha). \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

When we assume that the classical model is applicable we can insert in equation (10) the values of μ and α given by equation (7):

$$n_{+,-}^2 = \frac{\gamma^2 B_c (H_c + 2\pi M_0 \sin^2 \vartheta) - \omega^2 \pm \{(2\pi\gamma^2 B_c M_0 \sin^2 \vartheta)^2 + (4\pi\gamma\omega M_0 \cos \vartheta)^2\}^{\frac{1}{2}}}{\gamma^2 H_c (H_c + 4\pi M_0 \sin^2 \vartheta) - \omega^2} \quad (13)$$

From the equation we conclude that, according to the classical model, the resonance frequency ω_0 for a propagating wave depends on the direction of propagation and is given by $\omega_0^2 = \gamma^2 H_c (H_c + 4\pi M_0 \sin^2 \vartheta)$. At this frequency the value of n^2 of only one of the two possible waves becomes infinite. In the limit $\omega \rightarrow \omega_0$, n_-^2 remains finite.

As a second application of the phenomenological method we shall calculate again the alternating magnetic moment induced in the small permanently magnetized ellipsoid considered in §1. The result will now be a function of the tensor components μ and α . We notice that the ζ -axis is identical with the z -axis everywhere in the ellipsoid and that $H_c = H_0 - N_z M_0$. Further, the assumption that the dimensions of the ellipsoid are small, *i. e.* small compared with the wavelength of the high frequency electromagnetic field outside as well as inside the material, enables us to treat the problem in a quasistatic way: we may use Maxwell's equations in which the term $c^{-1}d\mathbf{D}/dt$ has been omitted. It follows then that inside the material all fields are homogeneous fields and that the following relations hold for the periodic parts of \mathbf{H} and \mathbf{M} :

$$\left. \begin{aligned} h_x &= h_x^0 - N_x m_x, \\ h_y &= -N_y m_y. \end{aligned} \right\} \quad (14)$$

With the aid of equation (6) we find

$$\left. \begin{aligned} m_x [\{4\pi + (\mu - 1)N_x\} \{4\pi + (\mu - 1)N_y\} - \alpha^2 N_y N_x] \\ &= [(\mu - 1) \{4\pi + (\mu - 1)N_y\} - \alpha^2 N_y] h_x^0, \\ m_y [\{4\pi + (\mu - 1)N_x\} \{4\pi + (\mu - 1)N_y\} - \alpha^2 N_y N_x] &= i\alpha h_x^0. \end{aligned} \right\} \quad (15)$$

A simple calculation shows that we obtain Kittel's equations (equation (1)) when we insert the values of μ and α given by equation (7) in equation (15).

§4. Quantum Theory.

In this section it is our aim to give a quantum-mechanical description of ferromagnetic resonance. We shall study the dispersion of electromagnetic waves of not too high a frequency in a saturated ferromagnetic material and treat the problem in a way analogous to the theory of the dispersion of electromagnetic waves in a dense dielectric medium. A difficulty in the latter theory is the fact that it is not sufficient to consider only the first order perturbation of the dielectric substance by the primary electromagnetic wave, since the dispersion phenomenon in a dense medium

is actually largely determined by the action of the secondary scattered waves on the dielectric medium itself. The usual approach to the dispersion problem in a dense medium is a semi-classical one. First the quantum theory of the interaction of one atom with an electromagnetic wave is worked out and it is found that the atom acts as a centre of secondary, coherently scattered waves. The combined effect of the primary wave and the secondary waves scattered from the atoms of the dielectric substance is then calculated in a classical way in order to find the dispersion law.

This procedure, however, is not applicable to a ferromagnetic medium. The reason is obvious: one single atom of a ferromagnetic substance cannot even approximately represent the complicated state of affairs in a ferromagnetic material since ferromagnetism is essentially a co-operative phenomenon. The best thing to do, therefore, is to consider a specimen of the material that is sufficiently large to show all characteristics of ferromagnetism but is still very small as compared with the wavelength of the disturbing electromagnetic wave. The second assumption is important since it allows us to neglect all retardation effects in the interaction between the electromagnetic field and the material. If we can show that the result of a quantum mechanical calculation of the coherent scattering of an electromagnetic wave from such a small body is identical with the result derived from the classical model described in §1, it seems reasonable to assume that the dispersion of electromagnetic waves in a ferromagnetic medium is in general given by the considerations in the foregoing sections.

We shall first study the eigenstates of a magnetized ferromagnetic body of ellipsoidal shape (with principal axes in the x , y and z direction) in which N atoms are arranged in cubic lattice. We use Heisenberg's picture of ferromagnetism. The ground state of an isolated atom (k) of the substance is supposed to be a $(2S+1)$ -fold degenerate state, corresponding to the $(2S+1)$ different orientations of an atomic spin vector S with respect to the z -axis. The wave functions are $\psi(m_k)(m_k=S, S-1, \dots -S)$. The Hamiltonian is assumed to be

$$H_{op} = -S^{-2} \sum_{k,l}^N U_{kl} \mathbf{S}_k \mathbf{S}_l + \sum_{k,l}^N 2\beta_0^2 R_{kl}^{-5} \{ R_{kl}^2 \mathbf{S}_k \mathbf{S}_l - 3(\mathbf{S}_k \mathbf{R}_{kl})(\mathbf{S}_l \mathbf{R}_{kl}) \} - \sum_k 2\beta_0 S_k^{(z)} H_0. \quad (16)$$

The three sums describe respectively an isotropic exchange interaction between nearest neighbours (\mathbf{S}_k is the spin angular momentum operator in units \hbar), the magnetic dipole-dipole interaction (β_0 is a Bohr magneton) and the interaction with the external magnetic field H_0 in the z -direction.

When H_0 is sufficiently large and the temperature is sufficiently low only those stationary states will play a rôle for which there is nearly complete alignment of the spins. Holstein and Primakoff (1940) have developed a method for dealing with these states in an approximate way. Their method can be considered as an improved Bloch spin-wave treatment of ferromagnetism; the improvement consists in including the dipole-dipole terms. Though the method of Holstein and Primakoff is open to

criticism we believe that their theory is as good as any other theory of ferromagnetism at low temperatures. Therefore we shall base our considerations in this section on the method given in Holstein and Primakoff's paper.

The mathematical device in Holstein and Primakoff's method is the introduction of new operators a_k and a_k^* with the properties

$$\left. \begin{aligned} a_k(\psi m_k) &= m_k^{\frac{1}{2}}(\psi m_k - 1), \\ a_k^*(\psi m_k) &= (m_k + 1)^{\frac{1}{2}}(\psi m_k + 1). \end{aligned} \right\} \dots \dots \dots (17)$$

They are related with the S_k operators by

$$\left. \begin{aligned} S_k^{(x)} + iS_k^{(y)} &= (2S)^{\frac{1}{2}}(1 - a_k^* a_k / 2S)^{\frac{1}{2}} a_k, \\ S_k^{(x)} - iS_k^{(y)} &= (2S)^{\frac{1}{2}} a_k^* (1 - a_k^* a_k / 2S)^{\frac{1}{2}}, \\ S - S_k^{(z)} &= a_k^* a_k. \end{aligned} \right\} \dots \dots \dots (18)$$

The operators satisfy the commutation rule

$$a_k a_l^* - a_l^* a_k = \delta_{kl}. \dots \dots \dots (19)$$

After rewriting the Hamiltonian in terms of the new operators, Holstein and Primakoff assume that, in order to study the properties of a ferromagnetic at low temperatures, it is sufficient to retain in the Hamiltonian only those terms which do not contain products of more than two operators a and a^* . The argument is that the expectation values of the neglected terms are likely to be negligibly small at temperatures well below the Curie temperature. For a criticism of this procedure, which we shall adopt in this paper, we refer to the original paper of Holstein and Primakoff.

We shall not use the operators a_k and a_k^* but the operators p_k and q_k defined by

$$\left. \begin{aligned} a_k &= (p_k + iq_k)(S/2)^{\frac{1}{2}}, \\ a_k^* &= (p_k - iq_k)(S/2)^{\frac{1}{2}}. \end{aligned} \right\} \dots \dots \dots (20)$$

The operators satisfy the commutation rule

$$p_k q_l - q_l p_k = i\delta_{kl}/S. \dots \dots \dots (21)$$

In the approximation of Holstein and Primakoff the Hamiltonian is in terms of the p and q operators :

$$\begin{aligned} H_{0p} = & C - \sum_{k,l} 12\beta_0^2 S^2 R^{-5} z(xp_l + yq_l) \\ & + \sum_{k,l} (12\beta_0^2 S^2 R^{-3} - U_{kl})(p_k p_l + q_k q_l - p_k^2 - q_k^2) \\ & - \sum_{k,l} 6\beta_0^2 S^2 R^{-5} \{x^2 p_k p_l + y^2 q_k q_l - z^2(p_k^2 + q_k^2) + 2xyp_k q_l\} \\ & + \sum_k \beta_0 S H_0(p_k^2 + q_k^2) \equiv C + G(p_k, q_l). \end{aligned} \dots \dots \dots (22)$$

Here C is a constant which does not contain p_k or q_k . Further, we have written x, y, z , and R instead of x_{kl}, y_{kl}, z_{kl} and R_{kl} .

For this we anticipate again the results of §6 which show that one of the operators c_i in (24) is given by

$$c_1 = \left(\frac{S\beta_0}{NF_1} \right)^{\frac{1}{2}} \sum_k [\{H_0 + (N_x - N_z)M_0\}^{\frac{1}{2}} p_k + i\{H_0 + (N_y - N_z)M_0\}^{\frac{1}{2}} q_k], \quad (25)$$

where

$$F_1 = 2\beta_0 \{H_0 + (N_x - N_z)M_0\}^{\frac{1}{2}} \{H_0 + (N_y - N_z)M_0\}^{\frac{1}{2}}. \quad (26)$$

With the aid of equations (18) and (20) it follows that

$$\left. \begin{aligned} Q_x &= \sum_k 2\beta_0 S_k^{(x)} = \left\{ \frac{NS\beta_0 F_1}{H_0 + (N_x - N_z)M_0} \right\}^{\frac{1}{2}} (c_1 + c_1^*), \\ Q_y &= \sum_k 2\beta_0 S_k^{(y)} = -i \left\{ \frac{NS\beta_0 F_1}{H_0 + (N_y - N_z)M_0} \right\}^{\frac{1}{2}} (c_1 - c_1^*). \end{aligned} \right\} \quad (27)$$

Here we have neglected terms containing products of more than two c -operators.

We suppose that \mathbf{h}^0 is directed along the x -axis and is given by $h_x^0 \cos \omega t$ and we assume that the ferromagnetic body is in a quantum state (k). By means of a method used in the semi-classical theory of coherent scattering (Kramers and Heisenberg 1925), we easily find that the body acts as a centre of magnetic dipole radiation, corresponding to the induced alternating magnetic moment

$$\begin{aligned} \Pi_{(k)} &= \sum_l \frac{-\frac{1}{2}h_x^0}{\hbar(\omega_{kl} - \omega)} (Q_{lk}^{(x)} Q_{kl} e^{i\omega t} + Q_{kl}^{(x)} Q_{lk} e^{-i\omega t}) \\ &+ \sum_l \frac{-\frac{1}{2}h_x^0}{\hbar(\omega_{kl} + \omega)} (Q_{lk}^{(x)} Q_{kl} e^{-i\omega t} + Q_{kl}^{(x)} Q_{lk} e^{i\omega t}). \end{aligned} \quad (28)$$

Since the Q operators involve c_1 operators we need only the matrix elements

$$c_1(n_1 | n_1 + 1) - (n_1 + 1)^{\frac{1}{2}} = c_1^*(n_1 + 1 | n_1), \quad (29)$$

and the energy difference

$$\hbar\omega_{n_1+1, n_1} = F_1. \quad (30)$$

With the aid of the foregoing equations and the relation $M_0 V = 2\beta_0 S N$ (V is the volume of the body) we find that for every quantum state of the ferromagnetic body

$$\left. \begin{aligned} \Pi^{(x)}/V &= \text{real part of } m_x, \\ \Pi^{(y)}/V &= \text{real part of } m_y, \end{aligned} \right\} \quad (31)$$

where m_x and m_y are given by the equations (1) if we put

$$\gamma = 2\beta_0/\hbar. \quad (32)$$

§5. Discussion.

In §4 we have shown that Kittel's formula (*cf.* equation (1)) can be regarded as a Kramers-Heisenberg dispersion formula. The resonance frequency ω_1 corresponds to a "spectral line" in the micro-wave spectrum of the ferromagnetic substance. ω_1 is determined by the relation $\hbar\omega_1 = F_1$, where F_1 is the energy difference between any two stationary states of the system, between which matrix elements of the total magnetic moment perpendicular to the constant field exist. The fact that the quantum mechanical calculations do not lead to any result that is essentially different from those derived from classical considerations is not surprising in view of the close analogy which appears to exist between the classical model and Holstein and Primakoff's procedure for dealing with the energy levels of a ferromagnetic substance. Nevertheless the quantum mechanical approach may have its advantages for the considerations on the following two points: the deviation of the experimental value of γ from that expected for a free spin and the shape of the resonance curve.

We shall discuss the second point first. One can expect that the "spectral line" will not be a sharp line for several reasons. First of all it will have a natural width, which, however, will be extremely small. Secondly, we have given only an approximate description of the quantum states in a ferromagnetic body, since we have ignored in the Hamiltonian a number of terms which contain products of more than two c -operators. The terms describe the interaction between the different spinwaves (i). They are responsible for the establishment of thermodynamic equilibrium in the system of spinwaves and they will give rise to a broadening of the spectral line. Thirdly, there will be an interaction between the spinwaves and the lattice vibrations, which is responsible for the establishment of thermodynamic equilibrium between the lattice and the spin system. This interaction will also broaden the spectral line. An estimate of the times required to establish equilibrium by means of the two mechanisms has been made by Akhieser (1946). It must be pointed out, however, that Akhieser's calculations apply to the bulk of the spinwaves (i), which have wavelengths that are not small in comparison with the dimensions of the body, but which are so small that the exchange interaction is the predominant factor determining the energy of the spinwaves. (Wavelengths $< 10^{-5}$ cm., *cf.* §2). The spinwaves (i) which are involved in the resonance phenomenon do not belong to the bulk. Therefore we believe that the times calculated by Akhieser cannot be used to obtain an estimate of the broadening of the resonance curve. Preliminary calculations on this point have been made and they suggest that at sufficiently low temperatures and sufficiently high fields the resonance curve might be much sharper than it would follow from the relaxation times calculated by Akhieser.

As to the value of the resonance frequency, the quantum theory given in this paper leads to the equation $\hbar\omega_1 = F_1 = \beta_0 f(H_0)$. Here F_1 is the spacing of the energy levels involved; F_1 depends on the magnetic

Here G is the function defined in equation (22), when we put

$$\beta = 2\beta_0 S. \quad (35)$$

C'' is a constant which does not contain ξ_k or η_k . Since ϵ has a minimum value when all ξ_k and η_k are zero, the terms in ϵ which are linear in ξ or η must vanish (neglecting deviations for surface atoms). From now onwards we shall omit these terms and the corresponding terms in equation (22), which are linear in p or q .

When we confine ourselves to small ξ_k and η_k , the classical equation of motion for a top ($d\mathbf{J}_k/dt = \mathbf{T}_k$) can be written as

$$\left. \begin{aligned} J d\xi_k/dt &= \partial\epsilon/\partial\eta_k, \\ -J d\eta_k/dt &= \partial\epsilon/\partial\xi_k. \end{aligned} \right\} \quad (36)$$

At this point we want to show the correspondence between the classical and the quantum mechanical equations of motion. The latter are given by

$$\left. \begin{aligned} dp_k/dt &= i\hbar^{-1}(H_{op}p_k - p_k H_{op}), \\ dq_k/dt &= i\hbar^{-1}(H_{op}q_k - q_k H_{op}). \end{aligned} \right\} \quad (37)$$

In virtue of the equations (21) and (22), equations (37) can be written as

$$\left. \begin{aligned} dp_k/dt &= (\hbar S)^{-1} \partial H_{op} / \partial q_k, \\ -dq_k/dt &= (\hbar S)^{-1} \partial H_{op} / \partial p_k. \end{aligned} \right\} \quad (38)$$

Taking into account the correspondence between H_{op} in (22) and ϵ in (34) the correspondence between (36) and (38) is obvious.

We now resume the discussion of the classical problem. We seek a solution of the equations (36) for which all ξ_k and η_k are periodic functions of the time, *i. e.* proportional to $\exp(i\omega t)$. We write

$$G(\xi_k, \eta_l) = \frac{1}{2} \sum_{k,l} A_{kl} \xi_k \xi_l + \sum_{k,l} B_{kl} \xi_k \eta_l + \frac{1}{2} \sum_{k,l} C_{kl} \eta_k \eta_l \quad (39)$$

and observe that $A_{kl} = A_{lk}$, $B_{kl} = B_{lk}$, and $C_{kl} = C_{lk}$, that A_{kl} , B_{kl} and C_{kl} are real and that G is positive definite. The problem is to find the solutions of the following system of equations:

$$\left. \begin{aligned} \sum_l \{A_{kl} \xi_l + (B_{kl} + i\lambda \delta_{kl}) \eta_l\} &= 0, \\ \sum_l \{(B_{kl} - i\lambda \delta_{kl}) \xi_l + C_{kl} \eta_l\} &= 0, \end{aligned} \right\} \quad (40)$$

where $k = 1 \dots N$, and $\lambda = J\omega$.

Non-trivial solutions of the system of equations only exist when the determinant of the system is equal to zero. This condition leads to an equation of the N th degree in λ^2 . For the sake of simplicity we shall rule out the case of degeneracy by assuming that the roots of the equation are all different. They can be arranged in pairs in the following way:

$$\lambda_1, \lambda_2 = -\lambda_1, \quad \lambda_3, \lambda_4 = -\lambda_3, \dots, \lambda_{2N} = -\lambda_{2N-1}.$$

The values of ξ_k and η_k ($k=1 \dots N$) which satisfy (40) for a given value $\lambda=\lambda_i$ are $\xi_k^{(i)}$, $\eta_k^{(i)}$ ($k=1 \dots N$). The following equation can be derived from (40) :

$$\sum_{k,l} A_{kl} \xi_k^{(j)*} \xi_l^{(i)} + \sum_{k,l} B_{kl} (\xi_k^{(j)*} \eta_l^{(i)} + \eta_k^{(j)*} \xi_l^{(i)}) + \sum_{k,l} C_{kl} \eta_k^{(j)*} \eta_l^{(i)} + i\lambda_i \sum_k (\xi_k^{(j)*} \eta_k^{(i)} - \eta_k^{(j)*} \xi_k^{(i)}) = 0. \quad (41)$$

When we take into account the properties of A, B, and C we can deduce from the case $i=j$, that the eigenvalues λ_i are real quantities and different from zero. Further, we can derive that

$$\sum_k (\xi_k^{(i)*} \eta_k^{(i)} - \eta_k^{(i)*} \xi_k^{(i)}) \neq 0.$$

Therefore the eigenfunctions ($\xi_k^{(i)}$, $\eta_k^{(i)}$) can be normalized in such a way that

$$\sum_k (\xi_k^{(i)*} \eta_k^{(i)} - \eta_k^{(i)*} \xi_k^{(i)}) = \pm iS^{-1}. \quad (42)$$

(Observe that S is a real positive number.) When we assume that $\lambda_1, \lambda_3, \dots$ are the positive eigenvalues, the (+) sign corresponds to the cases $i=1, 3, 5, \dots$ and the (−) sign to $i=2, 4, 6, \dots$

When we interchange the indices i and j in equation (41) and take the complex conjugated values, we derive at once that if $i \neq j$:

$$\sum_k (\xi_k^{(j)*} \eta_k^{(i)} - \eta_k^{(j)*} \xi_k^{(i)}) = 0. \quad (43)$$

The equations (42) and (43) suffice to show that the eigenfunctions ($\xi_k^{(i)}$, $\eta_k^{(i)}$) ($i=1 \dots 2N$) are linearly independent. It follows that for a given eigenvalue λ_i there is only one eigenfunction. Finally, it is evident that

$$\xi_k^{(2n)*} = \xi_k^{(2n-1)}, \quad \eta_k^{(2n)*} = \eta_k^{(2n-1)} \quad (k=1 \dots N). \quad (44)$$

The foregoing analysis can be used to carry out the transformation of the Hamiltonian (22) into the Hamiltonian (24). We define the operators

$$c_i = \mp iS \sum_k (\eta_k^{(i)} p_k - \xi_k^{(i)} q_k). \quad (45)$$

Here the (−) sign should be taken for $i=1, 3, 5, \dots$ and the (+) sign for $i=2, 4, 6, \dots$; $c_{2n}^* = c_{2n-1}$. With the aid of (42) and (43) it is easily shown that the following commutation rules hold :

$$c_i c_j^* - c_j^* c_i = \pm \delta_{ij} \begin{pmatrix} + \text{sign for } i=1, 3, 5, \dots \\ - \text{sign for } i=2, 4, 6, \dots \end{pmatrix}. \quad (46)$$

With the aid of the inverse transformation

$$p_k = \sum_i \xi_k^{(i)*} c_i, \quad q_k = \sum_i \eta_k^{(i)*} c_i, \quad (47)$$

we find, taking into account equations (22), (39), (40), (42) and (43) :

$$H_{op} = C' + \sum_{i=1, 3, 5}^{2N-1} \frac{\lambda_i}{S} c_i^* c_i. \quad (48)$$

With regard to the equations (45), (47) and (48) it should be noticed that in the explicit expressions for $\xi_k^{(i)}$, $\eta_k^{(i)}$ and λ_i , the quantity β must be

replaced by $2\beta_0 S$ according to equation (35). Equation (48) is then identical with equation (24) if we put $\lambda_i = F_i S$.

From the considerations in §1 we know that there is one simple eigenprecession ($\xi_k^{(1)}, \eta_k^{(1)}$) of the classical system of tops, in which all atoms perform the same motions. The eigenfrequency ω_1 is equal to the resonance frequency in equation (1) :

$$\omega_1 = \beta J^{-1} \{H_0 + N_x - N_z\} M_0^{\frac{1}{2}} \{H_0 + (N_y - N_z) M_0\}^{\frac{1}{2}}. \quad (49)$$

With $F_i = J\omega/S$ and equation (35), the expression (26) follows immediately. The normalized eigenfunction itself is $\xi_k^{(1)}, \eta_k^{(1)}$ and is easily found from the considerations in §1 :

$$\left. \begin{aligned} \xi_k^{(1)} &= \left[\frac{\gamma}{2S\omega_1 N} \{H_0 + (N_y - N_z) M_0\} \right]^{\frac{1}{2}}, \\ \eta_k^{(1)} &= i \left[\frac{\gamma}{2S\omega_1 N} \{H_0 + (N_x - N_z) M_0\} \right]^{\frac{1}{2}}. \end{aligned} \right\} \dots \dots (50)$$

With the aid of equations (45), (49), (50) and (35) we find equation (25).

Finally, the author wants to express his sincere thanks to Prof. N. F. Mott for making available a research fellowship in this laboratory during the past year and for his hospitality.

References.

- AKHIESER, A., 1946, *J. Phys. U.S.S.R.*, **10**, 217.
 GRIFFITHS, J. H. E., 1946, *Nature, Lond.*, **158**, 670.
 HOLSTEIN, T., and PRIMAKOFF, H., 1947, *Phys. Rev.*, **58**, 1098.
 KITTEL, C., 1947, *Phys. Rev.*, **71**, 270 ; 1948, *Ibid.*, **73**, 155.
 KRAMERS, H. A., and HEISENBERG, W., 1925, *Z. Physik.*, **31**, 681.
 KRAMERS, H. A., and HELLER, G., 1934, *Proc. Acad. Amsterdam*, **37**, 378.
 LANDAU, L., and LIFSHITZ, E., 1935, *Phys. Z. Soviet Union*, **8**, 153.
 SNOEK, J. L., 1947, *Nature, Lond.*, **160**, 90.
 YAGER, W. A., and BOZORTH, R. M., 1947, *Phys. Rev.*, **72**, 80.

IX. Scattering of X-Rays by the L Electrons of Boron.

By KESSAR ALEXOPOULOS and HELEN SCOULOUDI *,

Physics Laboratory, University of Athens.

[Received March 1, 1948.]

SUMMARY.

In the present investigation the intensity of X-rays scattered by boron at different angles was measured. A monochromatic beam was obtained by the bent crystal method of reflection. The intensity of the scattered rays was measured in a scattering chamber of the Seemann-Bohlin type by the photographic method.

* Communicated by Prof. D. R. Hartree.

A comparison of the experimental results with the theoretical curves calculated for the case of zero, one and three free electrons per atom, shows that the experimental curve agrees best with the theoretical curve in the case of zero free electrons.

1. INTRODUCTION.

ONE of the ways to determine the number of free electrons in a solid is the scattering of X-rays by it. An investigation of this kind is possible because of the difference in the scattering power between free and bound electrons.

The intensity of the rays scattered by the *bound electrons* of an atom is given (Compton and Allison 1935) by the formula

$$I = I_e \left[\left(\sum_1^Z f_{nn} \right)^2 + R \left(Z - \sum_1^Z f_{nn}^2 - \sum_{n \neq k}^Z f_{nk}^2 + Z I_m \right) \right], \quad . \quad . \quad . \quad (1)$$

where I_e represents the intensity of the rays scattered by a free electron for the angle $\phi=0$, f_n is the electronic structure factor of the n th electron, Z is the number of electrons per atom and R is a correction due to the relativity effect (recoil factor). The first term of the formula (1) represents a coherent radiation, therefore a radiation liable to interference. The remaining terms give the incoherent radiation due to the finite dimensions of the atom. In formula (1) the last term $Z I_m$ which is a further relativity correction (Klein-Nishina correction) is negligible for ordinary wavelengths.

The radiation due to the scattering by a number of atoms is found by addition of the amplitudes scattered by each atom. When the atoms are arranged in a lattice then, owing to a destructive interference, the first term in the summation will be zero for all the scattering angles except for certain definite directions given by Bragg's law. For these, the interference is constructive, resulting in scattered rays of high intensity (*e. g.* Laue spots, Debye-Sherrer lines, etc.). The remaining terms, owing to their incoherent character, give no interference and therefore show a radiation continuously distributed in all directions, causing the continuous background of the Debye-Sherrer spectrograms. When considering atoms of a lattice in thermal agitation, the intensity of the first term according to Debye (1914) and Waller (1925) reduces to $I_e \left(\sum_1^Z f_n \right)^2 e^{-2M}$, where M is a function of the characteristic temperature of the crystal, while the intensity of the continuous background increases by the amount $I_e \left(\sum_1^Z f_n \right)^2 (1 - e^{-2M})$ * (Compton and Allison, *loc. cit.*, equ. 3.106). There-

* In view of the strong diffuse maxima discovered by Laval, Lonsdale and others, Debye's expression can no longer be considered correct for the diffuse scattering by single crystals. But as in the present case the scatterer is a crystalline powder Debye's expression can be considered a good approximation (Zachariasen 1940, Lonsdale 1942).

fore, for the case of an atom in a lattice, the intensity of the diffused scattered radiation is given by

$$I_{\text{diff}} = I_e \left[\left(\sum_1^z f_{nn} \right) (1 - e^{-2M}) + R \left(Z - \sum_1^z f_{nn}^2 - \sum_{n \neq k} f_{nk}^2 \right) \right] \quad (2)$$

As the quantities I_e , R , f_n , M can be calculated as a function of the scattering angle, the exact calculation of the radiation scattered by the bound electrons of an atom in a crystal becomes possible.

The case of the scattering of X-rays by *free electrons* was originally investigated by Zener (1935) who, by taking into account Pauli's exclusion principle, arrived at the following formula :

$$I = Z I_e \left[\frac{1}{2} (3x - x^3) \right], \quad (3)$$

where Z = the number of the scattering free electrons,

$x = 1/\lambda 2 (3n/\pi)^{-1/3} \sin \phi$,

λ = the wave-length of the radiation,

n = the number of free electrons per unit volume.

In fact, the electrons of the inner shells of each atom of the lattice are scarcely influenced by the neighbouring atoms and therefore have approximately the properties of the so-called bound electrons. On the other hand, the electrons of the outer shells are strongly influenced and therefore acquire properties which approach more or less those of the free electrons.

As this influence cannot be determined quantitatively, for the sake of comparing the experimental results with the theory, we make the following simplifying assumption. Some of the electrons of the outer shell behave as the corresponding electrons of the free atom (bound electrons), while the rest have properties of free electrons.

The intensity of the scattered radiation which is calculated according to the above approximation becomes :

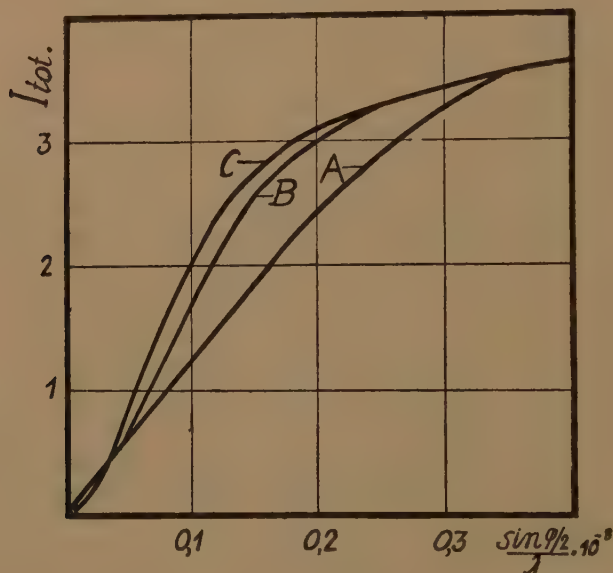
$$I_{\text{total}} = I_{\text{bound electr.}} + I_{\text{free electr.}} \quad (4)$$

where the intensities I_{bound} and I_{free} are given by the formulæ (1) and (3). The distribution of the continuous background which results from equation (4) depends upon the number of electrons which are considered as free. By comparing the distribution derived in this way with the experimental distribution curve, the number of free electrons can be determined.

This method was followed by Scharwächter (1937), who found that for Be, in the range of $\frac{\sin \phi/2}{\lambda} \cdot 10^{-8} = 0.04 - 0.6$, the distribution of the continuous background agrees with the theoretical curve only on the assumption that two of the electrons of this element are free. This result agrees with the number which is concluded from chemical or other experiments, so that it becomes evident that the investigation of the angular distribution of the continuous background can be considered as a fruitful method for investigating problems of this kind.

In the present paper the number of free electrons of boron was investigated (Scouloudi 1942). In this element two of its five electrons belong to the inner shell and therefore can be considered as bound. What fraction of the remaining three electrons we should consider as free will be decided upon by comparing the experimental results with the theoretical curves made under different assumptions. In fig. 1, the curves represent the distribution of the continuous background calculated for three, one and zero free electrons, the rest being treated as bound. These curves were calculated according to the method used by Scharwächter, by using equations (2), (3) and (4). The f_n values were interpolated from the data given by James and Brindley (1931) as based on Hartree's theory. In

Fig. 1.



Theoretical curves. A, B, C, cases of three, one and zero free electrons.

calculating (Scouloudi 1942) the theoretical curves B and C, the extra negative Waller-Hartree term $\sum_{n \neq k} f_{nk}^2$ of equation (2), due to the Pauli's exclusion principle, has been omitted, as this term is negligible as a rule (Compton and Allison, *loc. cit.*)*. It could, in fact, not be calculated as the wave functions for the electrons of boron atoms are not known. For the characteristic temperature of boron the value 1530°K. (Magnus and Danz 1926) was used. It is evident that for the range of angles less than $\frac{\sin \phi/2}{\lambda} \cdot 10^{-8} = 0.35$ there is a remarkable difference between the

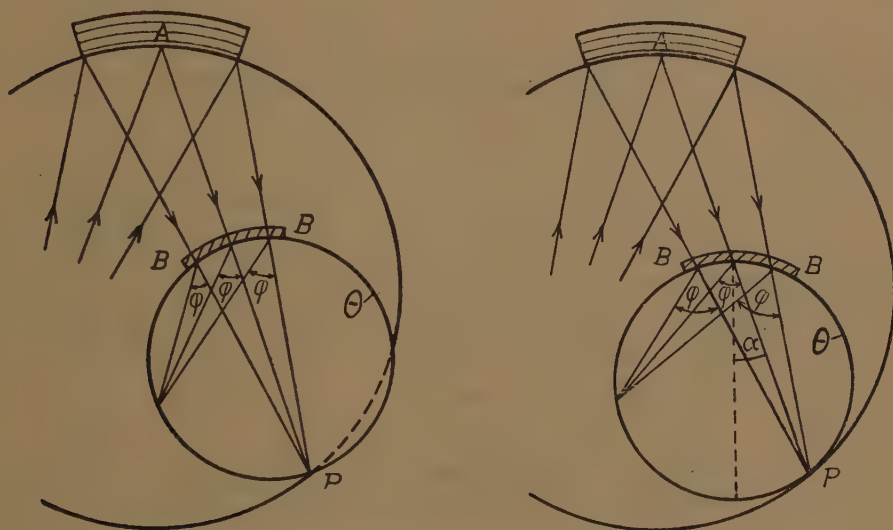
* It appears from work by Harvey, Williams and Jauncey⁽¹⁰⁾, Jauncey and Clans⁽¹¹⁾ and Jauncey and Denning⁽¹²⁾ on the diffuse scattering of X rays that, for crystals with neon-like atoms (NaF, KCl, MgO, SiC), this term is not negligible. A small inaccuracy may therefore be introduced in the following calculations by neglecting this term.

three curves, so that it should be possible to decide which of them agrees best with the experimental curve.

2. EXPERIMENTAL ARRANGEMENTS.

The distribution of the continuous background was found by scattering experiments of monochromatic X-rays upon a polycrystalline specimen of boron. The scattered intensity was measured by the photographic method. Because of the small atomic number of boron ($Z=5$), the scattering is very weak and therefore a monochromator capable of giving a very intense beam is needed. As such, a bent mica crystal A (lattice const. $d=19.9 \text{ \AA}$) was used (Guinier 1938) (fig. 2). The reflection took place on the cleavage face in the tenth order. The arrangements for bending the crystal ($R=440 \text{ mm.}$) and for adjusting follow the method used by Sandström (1935).

Fig. 2.



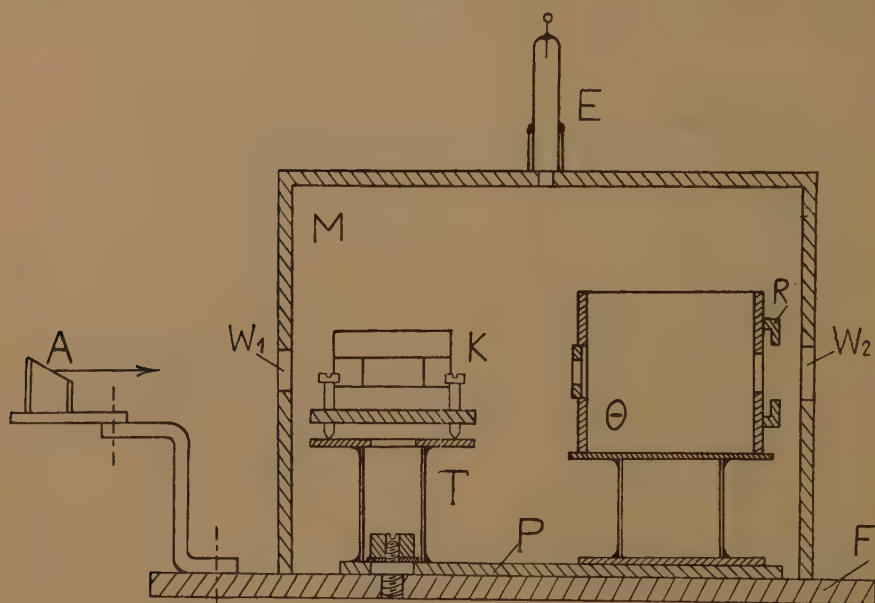
Symmetrical and unsymmetrical positions of the scattering chamber.

Because the monochromatic beam given by the bent crystal converged, a scattering chamber based on the principle of Seemann (1919) and Bohlin (1920) was used. On a short part BB of the chamber circumference the polycrystalline grains of boron were held between two thin films of collodium, in a layer 1.4 mm. thick. The photographic film was fastened on the rest of the circumference of the chamber. The diameter (20 mm.) of the chamber was chosen in such a manner that the place in which the scattering substance was held was about midway between crystal and film. This case, as Guinier⁽¹⁴⁾ showed, presents great advantages. Two positions of the scattering chamber with regard to the incident beam were used, according to the ranges of scattering angles for which the measurement was to be done ($\phi=6^\circ-75^\circ$): the *symmetrical*, for scattering angles ϕ of less than 45° , and the *unsymmetrical*, for angles ϕ greater than 45° .

To avoid additional scattering by the molecules of the air, the whole arrangement, *i. e.* the monochromator and the scattering chamber, was covered with an iron bell M (fig. 3) which had two windows W_1 , W_2 of celluloid foil for the entrance and exit of the primary beam. The bell was connected with a mercury diffusion pump.

The scattering substance.—Boron used in the present investigation was kindly given by Prof. Gudden, from Erlangen. It had a purity of 99.97 per cent being produced in crystalline state according to the method of Weintraub. Such a degree of purity was necessary because

Fig. 3.



Experimental arrangement.

A=target of the X-ray tube, K=monochromator, Θ =scattering chamber.

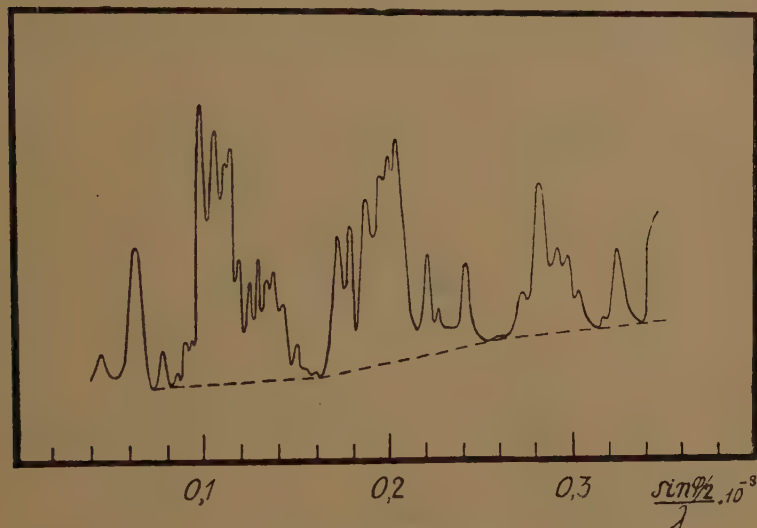
the least amount of foreign substances of greater atomic number would be sufficient to produce a blackening of the photographic films comparable with that of boron.

Exposures.—The exposures were made with the radiation from a Cu target working at a tension of 50 KV and a load of 20 mA. The films used were Agfa Röntgen super-special and the exposures lasted up to 60 hours. The films were enclosed in envelopes of special black paper which were fastened on the outer surface of the scattering chamber in suitable slits of the brass rings R (fig. 3). Each spectrogram presented a continuous background on which outstanding Debye-Scherrer lines could be observed.

3. DISCUSSION OF RESULTS.

Fig. 4, shows an example of a photometer curve of a film exposed for 60 hours *. The dotted curve on which the Debye-Scherrer lines are superposed represents the distribution of the diffused scattered radiation (continuous background). Owing to the great number of Debye-Scherrer lines, there is an uncertainty in drawing this curve, as there may be some doubt that the photometer curve actually reaches the background at all six points, as assumed in fig. 4.

Fig. 4.



Photometer curve.

On the basis of a comparison of the widths of the lines to the separation of adjoining maxima, however, we feel certain that the background is actually reached at the first five points of fig. 4 and from now on we shall restrict our results to these only. The sixth point, as it can be seen, corresponds to a scattering angle at which all three theoretical curves coincide, so that by neglecting it the discussion of the results will not be affected. For the sake of comparison with the theoretical curves, the relative values derived from fig. 4 were corrected (a) for the variable distance between photographic film and the boron sample, (b) for the polarization factor and (c) for relativity effects.

The results from different films, after correcting in this way, are given in fig. 5, together with the theoretical curves of fig. 1. In fig. 5, crosses give the mean values from three different films of the same exposure,

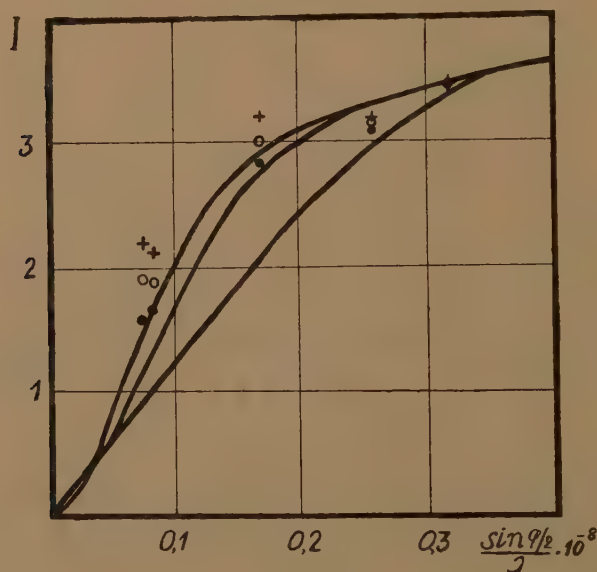
* The intensity of the Debye-Scherrer lines may not be derived from this figure, because at the peak values the blackening exceeded the graduated range of the photometric arrangement.

while dots are the results from another film exposed for a much longer time. The circles give the mean values for all the measurements made, the strongly exposed film being given a weight three times larger than the others. All values are reduced to such a scale that observations at the point $\frac{\sin \phi/2}{\lambda} = 0.318 \times 10^8$ coincide with the theoretical curves.

Although the results from the different films do not coincide, they all show the same trend and from the comparison with the theoretical curves it can be noticed that *the curve calculated for the case of zero free electrons appears to approach the experimental results most closely.*

The above results should be compared to the experimental data available for boron based on different methods of investigation.

Fig. 5.



Comparison of the experimental points with the theoretical curves.

The first experiments on the conductivity of boron gave the impression that this element is a semi-conductor, because it shows a high resistivity which reduces considerably as the temperature rises (Weintraub 1913, Freumann and Stieber 1934). Recently (Henniger 1937) however, boron specimens of low resistance have been produced by special methods. But even this good conductivity is far less than that expected for a metal.

On the other hand, the results obtained from the investigation of the fine structure of the lines of X-ray spectra are contradictory. O'Bryan and Skinner (1934) found that the boron K line is symmetrical, which agrees with the shape expected for an insulator. On the contrary, Hautot and Serpe (1937) found an unsymmetrical shape of the line just as expected for a conductor.

We wish to express our thanks to Prof. D. R. Hartree, F.R.S., for the interest he took in this subject and to Dr. W. Ehrenberg for his constructive criticism and discussions of the theoretical part.

REFERENCES.

- BOHLIN, 1920, *Ann. der Phys.*, **61**, 430.
- COMPTON, A. H., and ALLISON, S. K., 1935, *X-rays in Theory and Experiment*, 253-254.
- DEBYE, P., 1914, *Ann. d. Physik*, **43**, 49.
- FREUMANN, R. and STIEBER, A., 1934, *C.R. de l'Ac. de Science*, **109**, 1109.
- GUINIER, A., 1938, *Ann. de Phys.*, **12**, 161.
- HARVEY, G. G., WILLIAMS, P. S., and JAUNCEY, G. E. M., 1934, *Phys. Rev.*, **46**, 365.
- HAUTOT, H., and SERPE, J., 1937, *Journ. de Phys. et Ra.*, **8**, 175.
- HENNIGER, F. P., 1937, *Ann. der Phys.*, **28**, 245.
- JAMES R. W., and BRINDLEY, G. W., 1931, *Phil. Mag.*, **12**, 81.
- JAUNCEY, G. E. M., and CLAUS, 1934, *Phys. Rev.*, **46**, 941.
- JAUNCEY, G. E. M., and DEMING, 1935, *Phys. Rev.*, **48**, 577.
- LONSDALE, K., 1942, *Rep. on Progr. in Phys.*, **9**, 256.
- MAGNUS and DANZ, 1926, *Ann. der Phys.*, **81**, 407.
- O'BRYAN and SKINNER, H. W. B., 1934, *Phys. Rev.*, **45**, 370.
- SANDSTRÖM, A., 1935, *Dissertation, Uppsala*.
- SCHARWÄCHTER, W., 1937, *Phys. Zeits.*, **38**, 165.
- SCOULOUDI, H., 1942, *Thesis, Univ. of Athens*.
- SEEMANN, H., 1919, *Ann. der Phys.*, **59**, 455.
- WALLER, I., 1925, *Dissertation Uppsala*.
- WEINTRAUB, 1913, *J. Ind. Eng. Chem.*, **5**, 12.
- ZACHARIASEN, 1940, *Phys. Rev.*, **57**, 597.
- ZENER, C., 1935, *Phys. Rev.*, **48**, 573.

X. Notices of New Books and Periodicals received.

Existentialisme Théologique. Par ENRICO CASTELLI. [Pp. 96.] (Hermann & Cie, Paris, 1948.) No price.

EXISTENTIALISM is not a philosophy ; it is rather a trend or tendency inherent in a number of philosophies. It is almost what the Cynics termed a *βίος*, a way of life. It is especially influential to-day on the Continent, in Germany through Heidegger and Jaspers, and France through the novelist and dramatist Sartre and others. But its roots may be traced back to the strange and lonely Danish theologian Kierkegaard, and through him to earlier figures, notably Augustine.

The fundamental tenet of existentialism is that existence precedes being. Man, living in the present, is creating his future ; he is always in a process of becoming. At any given moment he is confronted with a choice ; because of the continual responsibility of choice upon him, man is in anguish. But by exercising his power of choice and by making personal decisions he can mould his future and free himself from this anguish. Such an outlook is anti-intellectualist ; it is " une philosophie de la catastrophe " (p. 16). It may be combined with a system of thought that is either theistic or atheistic ; it cannot fit one that is materialist or determinist. But it is a view of life which scientists cannot continue to ignore.

Of Prof. Castelli's book it suffices to say that if it won't do much good, at least it won't do any harm.

J. F.

Tables of Bessel Functions of Fractional Order.—Vol. I. Prepared by the Computation Laboratory of the National Applied Mathematics Laboratories, National Bureau of Standards. [Pp. xlii+418; 10½ ins. by 8 ins.] 1st Edition. (New York: Columbia University Press, 1948.) Price .

THE Computation Laboratory of the National Applied Mathematics Laboratories was organized in January 1938; but prior to July 1947 it has been called the Mathematical Tables Project, and since March 1943 it has been operated by the U.S. National Bureau of Standards. The present volume is another in the series of tables on Bessel functions prepared by what was then the Mathematical Tables Project, and previously reviewed in this Magazine (May 1944, April 1947, January 1948).

This volume tabulates $J_\nu(x)$ to 10 places of decimals for $\pm\nu=\frac{1}{4}, \frac{1}{3}, \frac{2}{3}, \frac{3}{4}$ and $x=0(0.001)0.9(0.01)25.0$. To extend the range to larger values of x , the Jacobi-Barnes asymptotic expansion

$$J_\nu(x) \sim A_\nu(x) \cos(x - \frac{1}{2}\pi\nu - \frac{1}{4}\pi) - B_\nu(x) \sin(x - \frac{1}{2}\pi\nu - \frac{1}{4}\pi)$$

is used; the functions tabulated being $A_\nu(x)$ and $B_\nu(x)$ to 10 places of decimals for the above values of ν and $x=25(0.1)50(1)500(10)5000(100)10000(200)30000$, $\sin x$ and $\cos x$ to 10 places of decimals for $x=0(0.01)1.6; 2(1)40$ radians, and $\frac{1}{2}\pi n$ to 15 places of decimals for $n=1(1)100$. For x -way interpolation when $x \geq 0.05$, J , A , and B are provided with second and sometimes fourth central differences, these differences being in some cases modified, with the object of attaining 10-place accuracy. Interpolation of this kind is not feasible for $|x| < 0.05$; so in this range $x \rightarrow J_\nu(x)$ is also tabulated with its second central difference. Tables are provided of the Everett coefficients of the second and fourth differences at intervals of 0.001. For ν -way interpolation

$$J_\mu(x) \simeq \sum_{\nu} L_{\nu}(\mu) J_{\nu}(x), \quad (\pm\nu = \frac{1}{4}, \frac{1}{3}, \frac{2}{3}, \frac{3}{4}),$$

the coefficients $L_\nu(\mu)$ of a seventh-order Lagrangian interpolation are given for $\mu=0(0.001)1$; and the relation $L_\nu(-\mu) = L_{-\nu}(\mu)$ extends this table to the corresponding negative values of μ . There is also a table of the first 30 zeros of $J_\nu(x)$ for $\pm\nu=\frac{1}{4}, \frac{1}{3}, \frac{2}{3}, \frac{3}{4}$ to 10 places of decimals; while for larger zeros there is a table of the appropriate coefficients in MacMahon's formula.

The forthcoming second volume will tabulate $L_\nu(x)$ in a similar fashion, except that it is proposed to quote $e^{-x}L_\nu(x)$ when $x > 25$.

The text is reproduced by a photo offset of the final typescript, which is quite satisfactory except in the table for $L_\nu(\mu)$ where the figures are a little too small and cramped for comfort.

A competent introduction surveys some of the uses of these functions in practical applications, their relationship with the Hankel functions and the Airy integral, and sundry recurrence formulæ which can be used to extend the range of ν . Amongst the immediate applications of these tables, one may note the approximate solution of the wave equation $y'' + p(t)y = 0$ in the neighbourhood of a zero of $p(t)$.

J. M. HAMMERSLEY.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XI. *The Magnetic Field of Massive Rotating Bodies* *.

By P. M. S. BLACKETT,
The University, Manchester †.

[Received October 29, 1948.]

I. INTRODUCTION.

IN a recent paper (Blackett 1947) the author has drawn attention to the approximate validity for the earth, the sun and 78 Virginis, of the relation

$$P = -\beta_1 G^{\frac{1}{2}} U / 2c \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

between the magnetic moment P and the angular momentum U , where G and c are the gravitational constant and velocity of light respectively and β_1 is a constant of the order of unity.

H. W. Babcock, whose measurements of the magnetic field of 78 Virginis (1947a) were the first made on any star, independently drew attention to the proportionality of P and U for these three bodies (1947b, 1947c). It was pointed out by the author that already in 1923, the validity of equation (1) had been implicitly recognized by H. A. Wilson, but that this had fallen into oblivion. A detailed discussion of the nature of the experimental evidence was also given by the writer, together with a survey of the main theories which had been put forward at various times to explain the origin of the magnetic field of the earth and other large rotating bodies. The conclusion was reached that it was improbable that the validity of the empirical relation (1) was accidental in origin and that therefore one must consider the possibility that it represented a general property of massive rotating bodies of roughly spherical shape. If this were indeed the case, then it followed that an explanation of it must be sought in a new fundamental property of matter not contained within the structure of present day physical theory. Moreover, it seemed likely that a full understanding of the effect was only likely to be achieved within the framework of a general theory embracing both gravitational and electro-magnetic phenomena. However, it does not seem likely that any of the unified field theories which have been put forward hitherto would be able to explain the effect. For the observed phenomena demand some essential asymmetry in nature, *e. g.* between positive and negative electric charges, whereas most field theories, at any rate, all such theories as deal with macroscopic quantities only, contain no such essential asymmetry.

* Communicated by the Author.

† This paper was prepared for, and read at the Eighth Solvay Conference in Brussels in October 1948, and is published with the agreement of the Solvay Institute.

Since in (1) the dipole moment is proportional to the angular momentum of the body, and since the latter quantity is the sum of contributions from the whole bulk of the body, it is hard to resist the conclusion that every part of the rotating body must contribute to the total magnetic field.

Now the only simple hypotheses in the form of a differential law which yields the integral relation (1) is the hypotheses of H. A. Wilson (1923), that a mass element moving with velocity \mathbf{v} produces a magnetic field at a distance r given by

$$\mathbf{H} = -\beta_1 \frac{G^{\frac{1}{2}}}{c} m \frac{\mathbf{v} \times \mathbf{r}}{r^3} \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

in analogy with the magnetic field of a moving charge.

It must be emphasized that equation (2) is certainly untrue if applied to free translating bodies, as it both gives magnetic fields which certainly do not exist and because it is inconsistent with the restricted principle of relativity. The writer pointed out, however, that in spite of these difficulties, it seemed useful to postulate the validity of (2) when applied to a mass element of a rotating rigid body, that is, when the velocity of a mass element is given by

$$\mathbf{v} = \boldsymbol{\omega} \times \mathbf{R},$$

when $\boldsymbol{\omega}$ is the angular velocity of the body and \mathbf{R} is the distance of the mass element from the centre of gravity.

Putting aside for the moment the obvious arbitrariness of this procedure, we can use (2) to calculate the external and internal field of given bodies and compare the results with experiment. No other simple expression consistent with the empirical relation (1) appears to exist. One must, however, bear in mind the possibility that ρ in (3) may possibly not be precisely equal to the local density, but might depend also to some extent on the nuclear constitution of the body, or on the local gravitational field.

An equivalent formulation of (2) is to state that a mass flux $\rho \mathbf{v}$, associated with a rotation, has the same magnetic field as that of a current density \mathbf{i} , given by

$$\mathbf{i} = -\beta_1 \frac{G^{\frac{1}{2}}}{c} \rho \mathbf{v}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

This is the relation used by Chapman (1948b) in calculating the field of the sun and other bodies.

For a spherical body in which the mass density is an axially symmetrical function it has been verified by Chapman (1948a) that the dipole component of the external field has the magnitude given by equation (1).

In this report, the argument of the previous work will be extended along the following main lines. In section II, an account will be given of recent experimental measurements of the magnetic field inside the

earth's crust. These experiments appear to offer the possibility of unambiguous proof that the outer layer of the crust does actually contribute to the external magnetic field of the earth in spite of the fact that it is the seat neither of electric currents nor electric charges, nor of magnetic materials of such magnitude and properties as to allow the effect to be explained within the framework of known physical laws. It follows that these experiments alone, quite independently of the questionable validity of the astronomical evidence, can, in principle, provide definite proof of a new property of matter. As yet, however, the results are still inconclusive.

In section III. the recent experimental measurements by Babcock of the magnetic field of certain stars will be described, and in section IV. a short review will be given of the evidence as to the nature of the sun's magnetic field and its supposed radial limitation.

In section V. the astronomical evidence is reviewed as a whole. Section VI. describes some recent work on the origin of the secular variation of the earth's field.

In section VII. two specific hypotheses are considered, which are both consistent with the present observations, but which predict different and, in principle, verifiable phenomena.

II. THE MAGNETIC FIELD INSIDE THE EARTH.

Dr. E. C. Bullard pointed out in a discussion that measurements in deep mines of the magnetic field of the earth, using standard survey instruments, should serve to distinguish between theories in which the crust of the earth above the place of observation does contribute to the magnetic field and those theories in which it does not. The former will be called bulk theories and the latter core theories.

We will consider here only the main component of the earth's field, that is, that part which corresponds to the field of a dipole situated at the centre of the earth and directed along the axis of rotation.

For a core theory, the vertical component V and the horizontal component H of the magnetic field will increase with depth, varying inversely as the cube of the distance from the centre. If d is the depth of observation and a the radius of the earth, then for $d \ll a$, the vertical and horizontal components will be given in terms of their values at the surface by the expression

$$V_d = V_0 \left(1 + \frac{3d}{a} \right), \quad H_d = H_0 \left(1 + \frac{3d}{a} \right). \quad (4), (5)$$

For a bulk theory the variation of one or both components will be different, according to the particular assumptions made as to the origin of the field.

Runcorn (Hales and Gough 1947) calculated the variation of V and H downwards, assuming that each mass element contributes a magnetic field given by (2), and found that V should increase approximately according

to (4) as for a core theory, but that H should *decrease* at about twice the rate. Chapman (1948b and 1948c), corrected an unnecessary approximation in Runcorn's derivation and gave the following more accurate expression for H

$$H_d = H_0[1 - 3(5\rho_1/k\rho - 1)d/a], \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where ρ_1 and ρ are the surface and mean densities of the earth and where k denotes the ratio I/I_0 of the moment of inertia I of the earth to the moment of inertia I_0 of a uniformly dense sphere of the same size and mass.

It is convenient to express the variation of V and H in the general forms

$$V_d = V_0 \left(1 + \frac{Ad}{a} \right), \quad H_d = H_0 \left(1 + \frac{Bd}{a} \right), \quad . \quad . \quad (7), (8)$$

where A and B are constants, characteristic for a particular theory, but independent of V_0 , H_0 and the magnetic latitude.

For all core theories we have from (4) and (5) that $A=B=3$.

For the bulk theory based on Wilson's hypothesis, we have $A=3$ from Runcorn's result, and from (6)

$$B = -3(5\rho_1/k\rho - 1). \quad . \quad . \quad . \quad . \quad . \quad (9)$$

For the earth, $\rho_1=2.8$, $\rho=5.5$ and $k=0.88$, giving $B=-5.7$. We see that this bulk theory gives a decrease of H downwards of nearly twice the magnitude of the increase for a core theory.

Other bulk theories will give different values for these constants. Suppose for instance that the earth is considered as a uniformly magnetized sphere. Then it is easily shown that V and H are the same at a small depth d as at the surface; in other words, for such a theory, $A=B=0$.

2.1. *Experimental measurements of V and H in Mines.*

The first experiments to determine the horizontal component of the earth's field below the surface of the earth were made at the suggestion of Dr. E. C. Bullard in a mine in South Africa by Hales and Gough (1947). From fifteen determinations made on three successive days, the difference ΔH between the horizontal field at the depth of 1463 metres and that at the surface was $(-25 \pm 4)\gamma$, where $\gamma=10^{-5}$ gauss.

Corrections had to be applied to this for the magnetic effect of several shale bands and two dykes in the area of observation. These were estimated to produce a decrease of H between 6 and 14γ , leaving a net decrease of H of between 11 and 19γ , giving $\Delta H = -(15 \pm 5)\gamma$.

Hales and Gough compared this result with the expected value of $+11\gamma$ for a core theory and -26γ for a bulk theory based on (3), using Runcorn's expression.

Further measurements were made by Runcorn (1948) and collaborators in a coal mine in Lancashire at a depth of 1240 metres. This mine is more suitable than the one in South Africa for such measurements as

the area is less disturbed magnetically, as shown by a ground magnetic survey supplemented by geological evidence. A magnetic survey along the mine gallery used was made to find a position free from the effects of local magnetic materials, steel rails, pit shafts, etc.

Two variometers were set up and compared at the surface. One was left at the surface and read at regular intervals to determine the diurnal variation. The other was taken down the mine where readings were made over a period of a few hours, after which it was brought back to the surface and its readings compared with the surface instrument to determine any change of zero. By comparing the run of the readings at the surface and below ground, the effect of the diurnal variation could be largely eliminated. Corrections for the temperature of the mine (43° C.) were then applied, using the instrumental temperature coefficient as determined in a separate experiment.

Two separate determinations of the change of V and H were made with the following results:

$$\Delta V = V_d - V_0 = +(25 \pm 10)\gamma,$$

$$\Delta H = H_d - H_0 = -(50 \pm 10)\gamma.$$

TABLE I.

Change of Horizontal Field Under-ground measured in units
of 10^{-5} gauss (1 gamma).

Place	Depth metres	V gauss	H gauss	ΔH obs. gamma	H(core)	H(bulk)
South Africa ($26^{\circ} 0' S. 28^{\circ} 0' E.$)	1463	0.29	0.15	$-(15 \pm 5)$	+14	-27
Lancashire						
	1240	0.44	0.17	$-(50 \pm 10)$	+10	-20
Mean	1350	0.37	0.16	$-(33 \pm 12)$	+12	-23

Putting in the values $d=1240$ metres and $a=6380$ KM in equation (4), which we have shown should hold both for a core theory and for the particular bulk theory under consideration, we find $\Delta V = +26\gamma$, in close agreement with the observations. This agreement between the observed and expected change of V provides a valuable check on the method of measurement and on the freedom of the place of observation from magnetic disturbances, and so gives one confidence in the measurement of the change in H .

In Table I. are set out the experimental data for the measurements of H in South Africa and Lancashire. In the last two columns are given the values to be expected on a core theory (equation (5)) and on the bulk theory (equation (6)).

It is seen that in both places the observed value of ΔH is negative and much nearer the value calculated for the bulk theory than for a core theory. In South Africa the observed value is about half the calculated value for the bulk theory and in Lancashire over twice as

great. Till the origin of this discrepancy is found the reliability of all the measurements must remain in some doubt. Possible causes are (a) the existence of undiscovered local magnetic anomalies, (b) change of zero of instruments during transit to and from mine, (c) incorrectly determined or irregular temperature coefficient, (d) difference in density of surface rocks*.

While awaiting further measurements and experimental check on all these points, it seems useful to take a crude mean of the two results, giving the figures in the last row of the table. Taken in this way, it is reasonable to claim that the measurements so far made give some evidence, even if not yet very strong evidence, against a core theory and in favour of the particular bulk theory under discussion.

The main objective of these and future experiments of this type is the precise determination of the coefficients A and B of equation (7) and (8) at different localities and depths, and under materials of different density and physical properties. On the bulk theory of the Wilson type, these coefficients should be independent of locality and depth, but vary with the surface density according to (9) †.

TABLE II.
Predicted and Observed Values of Coefficients A and B.

Coefficient	Core	Bulk (Wilson)	Bulk (uniform mag.)	Observations	
				Place	Obs. Value
A	+3.00	+3.00	0	Lancashire	+ 2.8
B	+3.00	-5.7	0	{ S. Africa	- 4.3
				{ Lancashire	-14.6
					} mean -9.9

It is convenient to collect together in Table II. the observed values of A and B and the values expected on the different theories.

This presentation of the data shows clearly how much better the observations agree with the fundamental bulk theory than with either a core theory or a bulk theory depending on the assumption of uniform magnetization.

2.2. *Can the observations be explained by the accepted laws of physics?*

Though it has been shown that the observations are roughly consistent with the postulate that each part of a rotating body contributes a magnetic field given by (2) or (3), it is necessary to enquire whether any alternative explanation for the observations can be found, using only the accepted laws of physics.

* In a recent set of measurements in a coal mine in Kent, V was found to decrease by 22y and H to increase by 30y. As explained above V would be expected to increase on any theory and so these observations must be tentatively ascribed to a magnetic anomaly. However, as yet no other evidence for the existence of such an anomaly has been found and therefore these results imply caution in the interpretation of the earlier ones.

† A somewhat different theory is outlined in section VII.

The only accepted ways in which the earth's crust could contribute to the magnetic field of the earth are

- (a) by being electrically charged so as to give a magnetic field by the rotation of the earth,
- (b) by being the seat of a system of real electric conduction currents,
- (c) by possessing a relatively high magnetic susceptibility.

The first possibility (a) can be ruled out immediately, as already at the time of Sutherland's work (1904), it was recognized that the charge density σ required ($\rho \approx 0.3G^{\frac{1}{2}}\rho$) would give rise to an electric field inside the earth of the order of 10^8 ev/cm. Such a field certainly does not exist.

On hypothesis (b) a system of circulating conduction currents would have to exist with a current density $i \approx 0.3G^{\frac{1}{2}}\rho v$, where v is the velocity of the part of the earth considered, due to the rotation of the earth. This implies a current density of about 3×10^{-9} amps. cm^{-2} near the equator. Though earth currents of a local and ephemeral nature do occur, it is quite certain that no such systematic circulation of current of this order of magnitude round the earth's axes can possibly exist. Since the resistivity of the earth's crust may vary from 10^5 ohm cm^{-2} for limestone to 10^{11} ohm cm^{-2} for granite or sandstone, one would observe potential differences along the E-W direction ranging from 10^2 to 10^8 volts per kilometre. It is quite certain that such potential differences do not occur.

Even if such currents did exist in the crust, their origin and maintenance could certainly not be found within the accepted laws of physics. It is only in the liquid core of the earth that the physical conditions for the maintenance of a net current circulating round the axes can be expected. However, attempts by Elsasser, Frenkel and others to derive such a current system from the convective motions and temperature differences in the liquid core have not proved markedly successful*. It is quite certain that there is no possible mechanism for such a system of real conduction currents in the rigid crust.

It is worth noting that it is easy to derive in quite a simple and direct way, the value of the current density i that must be postulated to exist in the crust to explain any given observed change of H downwards, as expressed by the corresponding value of the constant B . The expression found is

$$B = 3 - 4\pi ia/H_0, \quad \dots \dots \dots (10)$$

where H_0 is the surface field and a is the radius of the earth. Taking B (obs.) as about -6 , and giving H its value of 0.3 gauss for the equator, we find that $i = 3.6 \times 10^{-9}$ amps cm^{-2} , in agreement with the previous estimate.

To explain the observed decrease of H downward by the third possibility (c), that is, by the hypothesis that the crust itself is magnetic, requires that the intensity of magnetization of the earth is far higher than the

* See also section IV.

Babcock also found no detectable field for three stars of later spectral type, α Canis Majoris (F5), ϵ Pegasi (E0) and α Tauri (K5). One star, B.D.18°3789 was found to have a magnetic field which varied periodically between +7800 and -6500 gauss. This star, like all but one of the others showing magnetic fields, belong to a small class of A and early F stars called spectrum variables, in which the relative intensity of certain weak metallic lines varies periodically with the time. The period of magnetic variation of B.D.18°3789 is stated by Babcock to be the same as that of its spectrum variation, which is 9.25 days. One star β Cor.B. which shows a magnetic field but is not in the list of 20 spectrum variables given by Deutsch (1947), is nevertheless, stated by Babcock possibly to be one.

No mechanism to explain the property of spectrum variability has been put forward, though it is clear that such stars must undergo some type of mechanical and thermal oscillation. In one spectrum variable (α^2 Canum Venaticorum) periodic variations in velocity up to ± 10 K.M. per sec. have been observed, presumably associated with such an oscillation.

Babcock considers it likely that the spectrum variability and the possession of a magnetic field may be closely related properties of a star, and further, that the mechanism of underlying both may be related to that underlying the slow solar cycle of sunspot numbers and polarity and of the form of the Solar corona. These considerations led him to the view that the sun may be a magnetic variable, and that this might account for certain discrepancies between the measured field of the sun at different times (see section IV.).

Babcock argues that the striking variation of magnetic field observed or B.D.18°3789 makes it unlikely that the magnetic field of this star can have a fundamental origin, such as had been postulated by both him and by the writer from the observed proportionality of the magnetic moment and the angular momentum for the earth, the sun and 78 Virginis. For clearly the angular momentum of a star cannot change in the way required to explain the change in magnetic field. It could be argued, however, that a star, which did possess a magnetic dipole determined by its angular momentum, and which was in such a state of mechanical and thermal oscillation as to give rise to the phenomenon of spectral variability, would be also expected to undergo electromagnetic oscillations too. Moreover, all "non-fundamental" or "specific" theories of the origin of the field of astronomical bodies, such as those of Elsasser and Frenkel, ultimately relate the origin of the field to the rotation of a substantial portion of the body—in the case of the earth to the liquid core. Rapid changes in the magnitude and direction of the angular momentum of that *part* of a star to which is attributed its magnetic moment are almost as little likely to occur as are rapid changes in the angular momentum of the whole star. It appears, therefore, that the phenomenon of magnetic

variability requires the hypothesis of some kind of electro-magnet oscillation about a mean field, whichever origin, a fundamental or a specific one, is assumed for the mean field.

It will be noticed from Table III. that the mean fields of all the four stars for which a field has been established are of the order of 1000 gauss, though one of them shows a large oscillation about this value. It seems difficult to suppose that the origin of the field of a star with a constant magnetic field can be essentially different from the origin of the field of a magnetic variable; for if the two mechanisms were quite different, one would not expect roughly the same mean field. We will therefore assume the origin of the *mean* fields of stars with constant and variable magnetic fields are essentially the same, and we will use data for both types of stars to estimate the values of the constant β_1 in equations (1) and (3).

While awaiting further details of Babcock's measurements of the variable field of B.D.18°3789, we will assume that its mean field is $(7800-6500)/2=650$ gauss, as shown in Table III.

We will only consider stars of type A0 to F0 inclusive, since (a) no magnetic measurements have been made on earlier types, and (b) it is known that high rotational velocities disappear rather suddenly between the types F2 and F5.

To calculate β_1 , it is convenient to re-write the expression (18a) of the previous paper, by introducing the peripheral velocity $v=\omega R$, instead of ω , and replacing R and M by their values R_1 and M_1 expressed in terms of the sun's radius R_s and mass M_s . In this way we get

$$\beta = \frac{5}{2} \frac{c}{G^{\frac{1}{2}}} \frac{R_s^2}{M_s} \cdot \frac{R_1^2 H}{M_1 v} \cdot \dots \dots \dots (12)$$

The quantity k is the ratio I/I_0 of the moment of inertia of the star to that of a uniformly dense body of the same mass and radius, while η is a quantity introduced to take into account the non-uniform rotation of different parts of the star. It is defined as the ratio of the actual angular momentum of the star to that of a uniformly rotating body of the same size, mass and equatorial peripheral velocity. Inserting the values $G^{\frac{1}{2}}/c=8.62 \times 10^{-15}$ cm $^{\frac{1}{2}}$ gm $^{\frac{1}{2}}$, $R_s=6.97 \times 10^{10}$ cm., $M_s=2.0 \times 10^{33}$ gm. we get

$$\beta = \frac{706}{k\eta} \cdot \frac{R_1^2 H}{M_1 v} \cdot \dots \dots \dots (13)$$

Since it is not possible at present to measure H and v on the same star, it is necessary to use the mean value of v as determined for each spectral type. Westgate has given the frequency of occurrence of equatorial velocities for O and B and for A type stars. This data was reproduced in the former paper from a table compiled by Becker (1942). The mean velocity for O and B type stars is found to be 102 KM/sec. and 107 KM/sec. for A types. Since Westgate states that early F stars

have similar rotational velocities to O and B types, we will assume the former value for the F0 stars.

In columns 5 and 6 of Table III. are also given the values of radius and mass in terms of the sun for the spectral class as given by Becker.

For k we will assume the value 0.20 derived for the polytrope model* with $n=3$. For η , again as previously, we will provisionally assume the value of unity, though we will discuss the possible effect of non-uniform motion later.

The resulting values of β_1 as calculated from (13) are given in the last column. Since they are calculated assuming that each star has the mean velocity for its class, whereas we know that the actual velocities range at least from 25 to 250 KM/sec., the calculated values of β_1 should show a similar 10 to 1 dispersion. It will be of great interest to see whether further experimental observations do show a frequency distribution of magnetic fields similar to that of rotational velocities. With the data available, all that can be done is to take the crude mean of all the calculated values of β_1 , without regard to the inequality signs, as representing the best determinable value of β_1 for these stars. This is found to be 0.57. Assuming this mean value of β_1 as determined in this way to be an approximation to the true value of β_1 in equations (1) and (3), we can use the known frequency of given rotational velocities to calculate the expected frequency of occurrence of stars with given magnetic fields. Westgate's (1945) data shows that roughly one quarter of A type stars have peripheral velocities lying in each of the velocity ranges, 0 to 50, 50 to 100, 100 to 160, and 160 to 250 KM/sec. Using (13) to calculate H , we find that one would expect about one quarter of A stars to have magnetic fields lying in each of the ranges, 0 to 600, 600 to 1200, 1200 to 1900, and 1900 to 3000 gauss. With the present assumed minimum detectable field of about 500 gauss, one would therefore expect about a quarter of these stars to give no detectable field. Babcock observed 1 out of 5. We can conclude that Babcock's measurements are not inconsistent with the validity of relation (1), with $\beta_1=0.58$. Assuming this to be true we can calculate from (13) the mean fields of the spectral classes from B0 to F0 given in Table IV.

It is interesting to note that both the angular rotation and the mean magnetic field increases as one goes from A0 to F0 stars and has its maximum value for the latter type. The apparent sudden drop in rotation and so presumably in magnetic field sets in between types F2 and F5.

Part of the reason that the value of 0.58 for β_1 obtained here is markedly lower than the value 1.15 derived in the former paper from the data for 78 Vir. alone is that it now appears from Babcock's further measurements of the fields of other stars that 78 Vir. has probably a larger magnetic field than the average for its class and so is presumably rotating also faster than the average.

* Professor Freundlich has informed me that he considers this as the most probable value.

Part of the difference is also due to the assumption that $k=0.20$ instead of 0.16 as in the earlier paper. It is worth emphasizing that even if the relation (1) turns out essentially to have no general theoretical validity, it is very probable that within a given spectral class the magnetic field of a star is likely to be closely correlated with its angular momentum. So most of the analysis of this section may prove to be valid even if the fundamental explanation of the field is abandoned in favour of a specific theory; though of course in this case the constant β_1 would have to be given another interpretation.

Some recent theoretical work by Schwarzschild (1947) suggests that the interior parts of a star may be rotating much slower than the outside, leading to a value of η considerably less than unity, and so to correspondingly increased values of β_1 . Chapman (1948b) has shown that Schwarzschild's model for the sun gives $\eta \simeq \frac{1}{2}$ leading to values of β_1 of about 1.2. However, Schwarzschild's theory appears to lead to a

TABLE IV.

Mean expected Magnetic Field of Early Spectral Types.

Spectral Type	M_1 Mass	R_1 Radius	v Velocity KM/sec.	$\omega=v/R$ Ang. vel. rel. to sun	H magnetic field gauss
B.0	15	6.9	102	7.4	500
B.5	6	4.3	102	12	500
A.0	2.7	2.3	107	23	800
A.5	1.8	1.6	107	33	1200
F.0	1.5	1.4	102	36	1700
Sun G.0	1.0	1.0	2.0	1.0	30 *

larger decrease of rotational period at the poles of the sun compared with the value at the equator than is actually observed. So it is probable that η is not as low as Chapman calculates and therefore β_1 not as high.

Another source of possible error is in the values assumed for the mass and radius of stars of given spectral type. Dr. Babcock has pointed out to me that the stars with measured fields are peculiar stars, *e. g.* 78 Vir. is A2_p not A2, and that such stars may not have quite the same size and mass as the normal types. He mentioned that Deutsch estimates that these peculiar stars are probably about 1 magnitude brighter than the corresponding normal stars and are bluer, who considers for instance that an A0 star may have the same size and mass as a B8 star. Since from (13), β is proportional to R_1^2/M_1 , we see from Table IV. that the value of β_1 would be increased. However, Babcock also suggests that the

* See section IV.

greater absolute brightness of these stars may be due to the fact that they are probably appreciably flattened by rotation and so have a larger value of gravity at the poles than at the equator. Such a larger value would lead to increased brightness. One must remember that the only stars of a given type for which the magnetic field can at present be measured are that small fraction for which the axis of rotation is nearly in the line of sight. For only in such cases is the rotational broadening of the lines sufficiently small to allow the Zeeman effect to be measured. If, therefore, the polar region of rotationally flattened stars are brighter than the equatorial region, one will expect that stars for which the magnetic field can be measured will be brighter than normal, but will not necessarily be larger or heavier.

Similar considerations may possibly apply to the spectrum variables, though I have not found this mentioned in the literature. The spectral lines of spectrum variables are quite narrow, of the order of 1 or 2 A.U. broad, as can be seen in the spectra reproduced in the paper by Deutsch. This indicates a component of rotational broadening along the line of sight less than about 50 KM/sec. So these stars are either (a) oriented at random but with much small rotation than is normal for A type stars, or (b) have normal rotations but are oriented along the line of sight. According to the first hypothesis, the correlation of large measured magnetic fields with spectrum variability noted by Babcock, would imply an anti-correlation of large magnetic field with large rotation—a conclusion so improbable as to be rejected. Hence one concludes that the second hypothesis is correct, that is, that the spectrum variables have roughly the normal rotation, but have their axes along the lines of sight. It is not, however, necessary to assume that the property of spectrum variability is a special property only of the polar region of such stars. For the lines which show this property are generally not only sharp but rather weak, and so might well tend to escape detection in a star oriented at a large angle to the line of sight, due to the rotational broadening. If this is the correct interpretation the observed correlation of large magnetic fields with spectrum variability might be reformulated as a correlation between the *observability*, rather than the *existence* of magnetic fields and spectrum variability, both being dependent on absence of large rotational broadening due to approximate parallelism of the axis of rotation with the line of sight.

IV. THE SUN'S MAGNETIC FIELD.

The original measurements by Hale and his collaborators between 1912 and 1918 of the general magnetic field of the sun seemed to give convincing evidence that the field has a dipole character with a value at the pole of about 53 gauss. The method used was that of detection of slight traces of circular polarization in the wings of the lines. The

latitude variation was found to be that expected for a dipole and in addition a time variation with a period of $31\frac{1}{2}$ days, revealed by a variation of field at a given latitude, was detected which showed that the magnetic axis made an angle of about 6° with the mechanical axis.

In spite of the convincing nature of these results, considerable doubt has been expressed at various times as to the reality of the field. This doubt arose because a few of Hale's original observers failed to detect the effect. The measurements of most of the observers agreed, however, among themselves reasonably well.

In a recent discussion of the subject, Cowling (1945) has written: "The average observed separation of the Zeeman components is of the same order as the probable error of a single measurement. Because of its smallness, its reality is sometimes doubted. The present author sees no reason to share these doubts. To ensure that personal bias should not colour the results, the original observer measuring the Zeeman displacements was kept in ignorance of the hemisphere and latitude of the observations. Thus, if he were prone to imagine a non-existent phenomenon, he would give it sometimes one sign, sometimes the other, and in this aggregate his estimates would cancel out. Since, in spite of this, regular results were obtained, the phenomenon being observed was clearly not imaginary. In fact, though all observers have not been able to identify real Zeeman displacements, all those who have been able to identify them have found effects with the same sign.

"Moreover, no convincing explanation of the observations has been given, other than the existence of a magnetic field. The phenomenon requiring explanation is a difference in polarization between the two wings of a spectral line. Effects producing a pure displacement or pure broadening of the line cannot explain the polarization; no one has yet suggested a satisfactory alternative to the explanation in terms of a magnetic field."

Hale found that different spectral lines gave different values for the magnetic field. In general, weak lines gave the greatest field (about 50 gauss) and strong lines a weaker field (of the order of less than 10 gauss). Since it is generally assumed that strong lines originate at greater heights than weak ones, this was interpreted as indicating that the field fell off rapidly with height. Hale concluded that the field fell from 50 gauss to less than 10 gauss within a radial distance of some 300 KM. Cowling, however, doubts if this rapid decrease with height need be considered as real.

This supposed rapid falling off of field with height has been held by Rosseland and Chapman to show that the lines of magnetic force must be nearly horizontal. But this would give a variation of polarization across the disk quite different from that found by Hale. On the whole, it seems more likely that the failure to detect the Zeeman effect of strong lines must be in some way due to their greater width.

Further measurements have been made by Thiessen (1946) who used an interferometric method with a circular analyser. In a series of measurements in 1945 he found a magnetic field for the Fe 6173 line corresponding to a polar field of 53 ± 12 gauss. The variation with latitude and longitude were not measured, but he states that he verified many times that the circular polarization diminished to zero at the equator and the poles, as it should. However, quite recently *, 1947/48, Thiessen has repeated his measurements on the same line and has failed to find any field as large as 5 gauss! He states that the new measurements were more extensive than the old ones and probably more reliable. He points out, however, that the later measurements were made at sun spot maximum, whereas the earlier ones were some two years before this. He plans further measurements at the next sun spot minimum.

Very recently, H. D. Babcock (1948b) has reported a series of measurements of the sun's field using a Lummer Plate crossing a grating, together with an analyser for circularly polarized light. Among 42 sets of readings made between 1940 and 1947, magnetic fields between 6 and 60 gauss were found in 18, while the remaining 24 gave no measurable field or slight negative values. He concludes that "Hale's surmise that the field may be variable appears to be supported."

As has already been mentioned, Babcock suggested that a possible explanation of these discrepancies may be that the general magnetic field of the sun varies with the phase of the sun spot cycle, just as the field of B.D.18°3789 varies with the phase of its spectrum variability; moreover, he speculates that the fundamental mechanisms may be the same for these two stars, though of course, the magnitude of the field and its period of variation are very different. On this view, presumably the sun would be considered a spectrum variable too. And in a sense it must be, for since the spectrum of a sunspot differs from the normal solar spectrum owing to its lower temperature, the total average solar spectrum at sun spot maximum must differ, however slightly, from that at sun spot minimum.

Cowling surveys the various theories that have been proposed to account for the sun's field and finds that none of them give the right order of magnitude. The least unpalatable is that the magnetic field arises from electric currents resulting from the thermal motion of a rotating mass. This theory is analogous to that proposed by Elsasser to account for the earth's field. Cowling supposes that convective motion occurs in an unstable central region. The rising and falling masses are deflected to the east and west due to the Coriolis force. These Coriolis forces set up pressure differences in the gas which give rise to a relative diffusion of electrons and ions and so produce an electric current. Cowling shows that the direction of the magnetic field produced is that observed.

* Private communication.

However, by making reasonable assumptions as to the temperature differences, velocity of rise and fall, and electric conductivity, Cowling finds that the surface magnetic field of the order only of 10^{-6} gauss would be produced. So this theory must be abandoned.

Cowling can find no other theory which promises any better success, except to remark on the very unlikely possibility that the core of the sun might be capable of permanent magnetization. He writes: "This demands that hot ionized material is capable of acquiring a regular crystal-like structure, a possibility normally disregarded. In view of the difficulties of other hypotheses, the possibility may, however, be worthy of further study."

We can sum up the situation as follows:—(a) The sun probably possesses a dipole field whose polar strength may possibly fluctuate from about 50 gauss to less than 10 gauss, giving a mean field of about 30 gauss. (b) No explanation of the origin of the field has been found using only known properties of matter and none seem likely to be found.

It is of some interest to note that if the field of the sun is 53 gauss, then the values of β_1 , deduced from relation (1) is 0.90, whereas the mean value for the stars with measured fields is 0.58. If, however, the mean field of the sun over a sun spot cycle is in fact only about 30 gauss, one obtains a value of 0.52 that is closely the same as that obtained for the stars.

Three other methods exist by which information about the magnitude of the sun's field outside the chromosphere could be obtained, in principle; these are the form of the solar corona and of comets tails and the effect of the sun's field on the intensity of cosmic rays at the earth*. But no reliable information appears to be obtainable as yet from these phenomena.

Even if the sun's general field is supposed to have a fundamental origin, a quite different explanation must be found for the supposed obliquity of the magnetic axis and for sunspot fields. Presumably the former would be explained by postulating subsidiary induced dipoles as in the theory of the earth's secular variation developed by Elsasser and Bullard (section VI.).

It was recognized already long ago by Sutherland, and has recently been re-emphasized by Chapman, that the field of a sunspot is vastly larger than would be expected if it had the same fundamental origin as the main field. No fully satisfactory theory seems to have been found to explain the magnetism of sunspots, and it could be argued that when one is found, the same mechanism might serve to explain the main field. It is possible that this may prove to be the case, but on the other hand, it is also possible that the explanation of the field of a sunspot may lie in a drastic distortion of the main field, through some complicated hydrodynamic motions, such as have been discussed by Alfven, Elsasser and by

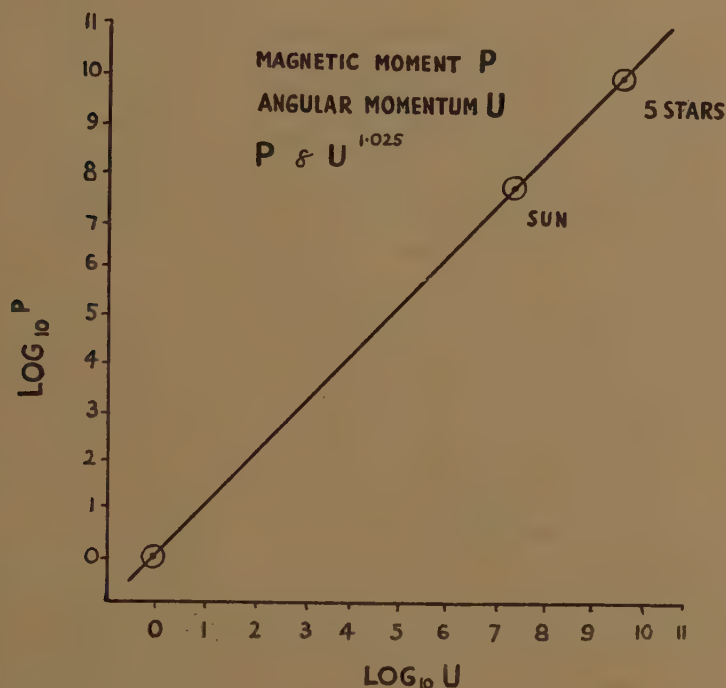
* See recent note by Alfven (1947).

Cowling. Elasser and Cowling (1933), in fact, appear to conclude from their analyses of such motions that their main effect is to compress or expand existing lines of force—the freezing in effect—rather than to create new ones, and so do not provide a theory of the main field, but might possibly provide one of sunspots provided the existence of the main field is assumed.

V. COLLECTED RESULTS.

From the data of Table III. and of Table IV. of the previous paper it can be deduced that the ratio of the mean angular momentum of the

Fig. 1.

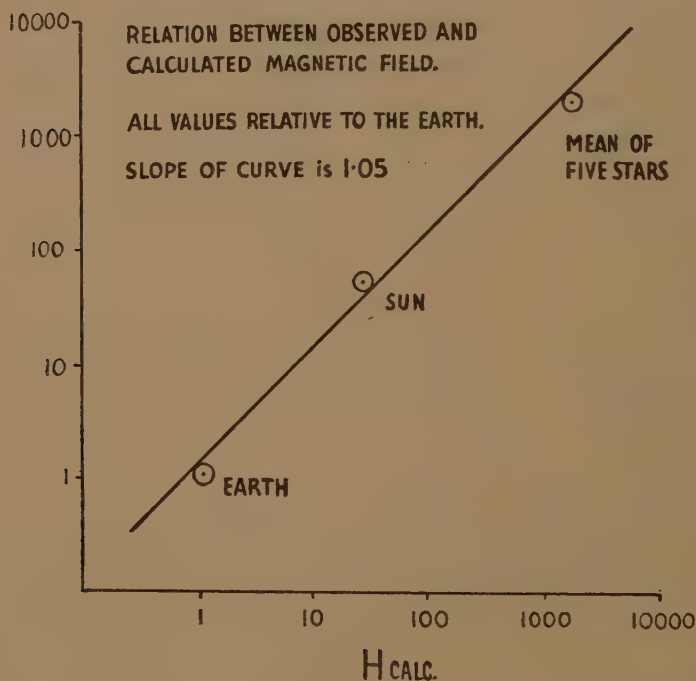


five stars to that of the sun, is 223 ± 50 . Taking the polar field of the sun as 30 gauss, the ratio of the mean magnetic moment of the five stars to that of the sun is found to be 230 ± 40 . So we see that for the mean of these five stars compared with the sun, the magnetic moments are proportional to the angular momenta within the experimental error.

In fig. 1 are shown the values of $\log_{10} U$ and $\log_{10} P$ for the earth, the sun and the mean of the five stars. The value of k for the earth is taken as 0.88, and for the sun and the stars as 0.20. The slope of the straight line drawn through the three points is 1.025, that is, unity within the experimental error. In these calculations η has been taken as unity, that is, all the bodies have been assumed to be in uniform rotation.

An alternative method of displaying the results is shown in fig. 2, in which the observed field H_{obs} relative to that of the earth is plotted on a log-log scale against the calculated field. This latter is calculated from the expression $H \propto M\omega/R$ which is derived directly from equation (1). This presentation of the results has the advantage over that shown in fig. 1 of comparing directly the measured field with the calculated field. The range of measured and calculated fields is about 1800 to 1, whereas the range of P and U is about 10^{10} to 1. The much greater range of the latter quantity arises because $U \propto M\omega R^2$ and $P \propto HR^3$, and R for the sun and the stars is about 100 times greater than for the earth.

Fig. 2.



VI. THE SECULAR CHANGE IN THE EARTH'S FIELD.

The situation outlined in the former paper, that no plausible theory exists of the earth's main field relying only on the known laws of physics, still persists. However, some progress has been recently made by Elsasser (1941, 1946) and Bullard (1948), in explaining the origin of its large secular variation, which may amount to 25 per cent of its average value.

It has long been recognized that the secular variation appears to be a regional rather than an earth wide phenomena (see for instance Chapman and Bartels, 1, 130). Elsasser showed by a statistical analysis of the spherical harmonic components of the earth's fields that the non-dipole

part of the field can be represented by some 10 dipoles distributed near the outside of the liquid core. On this interpretation the obliquity of the magnetic axis would seem to be due to a number of essentially local disturbances rather than to a systematic earth-wide phenomenon.

The main conception of Bullard and Elsasser is based on the observed fact that the field of the earth varies in a related way over certain large areas of the earth, of the order of a thousand miles or so in diameter. The period of variation, of the order of 100 years, is far too rapid to suppose that the cause lies in any thermal or mechanical changes in the solid outer 3000 KM of the earth. It is, therefore, necessary to seek the origin of the changing field in the liquid core of the earth, for only here can mechanical and thermal motions take place quickly enough.

Bullard shows that the observed secular change of field in South Africa over the last 100 years is consistent with the slow growth of a magnetic dipole near the interface between core and crust. He shows quantitatively how such a dipole can arise from a slowly growing eddy in the outer region of the liquid core, provided one assumes the pre-existence of the main field. Bullard calculates that a mass of electrically conducting fluid rotating in the main field will give rise to electrical eddy currents and so to an induced magnetic dipole, of such a character as to give an additional field at the surface of the type observed in a secular change. Making plausible numerical estimates of the size of the eddies, Bullard can explain at least the order of magnitude of the observed changes. Though there are still difficulties to be overcome before such explanation can be made quantitatively satisfactory, it does for the first time offer a plausible and already semi-quantitative origin of the secular variation, provided, and only provided, one assumes the pre-existence of a main field.

Bullard points out that his theory of the secular variation may provide, when more fully worked out, a test between a core theory and a bulk theory of the main field. For the magnitude and direction of the magnetic dipole induced in a rotating eddy in the core will depend on the magnitude and direction of the main field near the surface of the core, and this will be markedly different according to whether the main field itself arises in the core or whether it originates in the whole bulk of the earth. So far, it is not possible to judge which origin is indicated. In fact, neither theory at present seems adequate.

From this work of Bullard we can conclude that a plausible theory of the local secular variations may possibly be found by supposing that they are due to the interaction of the motions of the liquid core with the main field, but that the main field itself is not likely to be so explained, and so must rather be assumed to arise in some other way. The partial success of Bullard's theory of the secular variation gives, therefore, some indirect support for the view that the main field itself does not arise in the core. Now it is only in the liquid core that the main field can conceivably arise, provided the normal laws of physics are assumed. For there is almost certainly no possible physical mechanism in the solid outer part which

could give rise to the main field by known physical principles, *e. g.* currents, charges or permanent magnetism. So if it is found possible to prove that the main field does not arise in the liquid core, then we conclude that it cannot be explained at all by the accepted laws of physics, and so must arise from some new property of matter.

VII. DISCUSSION.

The experimental evidence discussed in the former sections is certainly not as yet adequate to prove conclusively that the earth's magnetic field arises from a new property of matter. Further measurements of the field of the sun and the stars and of the variation of the earth's field below the surface are clearly urgently needed. However, the existing evidence is sufficient to justify a careful search for new possibilities of testing the hypothesis of a fundamental origin of the field. One such possible test arises out of the following considerations.

We will start by assuming the approximate validity of equation (3), that is, that a mass flux ρv , associated with a rotation ω , has the same magnetic field as that of a current density i given by

$$i = -\beta_1 \frac{G^{\dagger}}{c} \rho v, \quad (3)$$

where $v = \omega \times R$. It will be convenient to call this current i , the virtual rotational electric current, or more simply, the virtual current associated with a rotational mass flux.

In seeking to find an explanation of this virtual current, two different theoretical possibilities present themselves in the first instance for consideration and lead to significantly different predictions.

The first is that equation (3) is to be taken as exact, that is, that ρ is to be taken as simply the mass density, that is, the sum of the masses of all protons, neutrons and electrons in unit volume. On this hypothesis, the constant β_1 , as determined experimentally either from the external or internal field of a rotating body, should always be found to have the same value. In fact, β_1 which from (3) is proportional to $i/\rho v$, and so can be considered as the virtual current per unit mass flux, would be the same for all bodies of whatever nuclear, chemical or physical constitution.

The second hypothesis is that the virtual electric current may depend not on the total mass of all the particles, but on their total electric charge, that is, may depend on the total number of electrons (or protons) in unit volume. Since the ratio of number of electrons to number of nucleons varies only by a factor of a little over two between heavy or light elements, this hypothesis is consistent with the *approximate* validity of (3). However, this variation is, in principle, detectable by the difference in the magnetic field produced by the rotation of matter of different nuclear constitutions. The difference arises, of course, simply from the fact that the neutrons, on this hypothesis, contribute to the mass but not to the virtual currents, so that for instance, a rotating mass of hydrogen

would have a larger magnetic effect per unit mass than a mass of heavier elements*.

For instance this second hypothesis can be tested in principle by determining the magnetic field of astronomical bodies with different nuclear constitutions (*e.g.* the earth and the sun, or two stars of markedly different hydrogen content), or (*b*) by measurements of the vertical variation of magnetic field of the earth in mines and under the ocean.

This second hypothesis would be in conformity with the view that the magnetic effect of rotating matter is in some way bound up with some minute inequality in the behaviour of positive and negative charges.

We see then that our two hypotheses satisfy one of the main requirements of useful hypotheses, in that they suggest further experiments. Since the second one has the more interesting consequences it is worth while considering it in more detail and formulating it explicitly. To do this we replace the mass density ρ in (3) by $(M/e)\sigma_0$, where σ_0 is the charge density of electrons (or protons), and e/M is the ratio of the charge to the mass of the proton. We will define σ_0 as essentially a positive quantity, just as is ρ in (3). Then we obtain

$$i = -\beta_2 \frac{M}{e} \frac{G^{\frac{1}{2}}}{c} \sigma_0 v, \quad . \quad . \quad . \quad . \quad . \quad (14)$$

where β_2 is a new constant, again of the order of unity.

We can, in principle, distinguish experimentally between our two hypotheses by finding which of the two equations (3) or (14) gives the same value of the constants β_1 and β_2 for bodies of different nuclear constitution.

From these equations we have

$$f = \frac{\beta_1}{\beta_2} = \frac{\sigma_0}{\rho} = \frac{\Sigma W}{\Sigma Z},$$

where Z and W are the atomic numbers and atomic weights of all the component nuclei. In Table V. are given the values of f for certain elements and mixtures of elements. If (14) is correct, but if we use (3) for convenience to calculate β_1 , then the relative values as determined experimentally for any two materials will be given by the corresponding values of f . If we compare the value of f for limestone ($f=0.50$) and water ($f=0.56$), we see that it is not completely excluded that measurements of the field underground and underwater might be made of sufficient accuracy to test between the two hypotheses. If the same value of β_1 is obtained, the first hypothesis is indicated; if β_1 for water is found to be 11 per cent higher than for the crust, then the second is indicated.

Of greater immediate interest, however, is the calculated difference between the value of f for the earth, assumed mainly of composition similar to olivine ($f=0.48$), and that for the probable constituents of the sun, *i.e.* 35 per cent of hydrogen by weight ($f=0.68$). The ratio

* The neutron here is taken as an elementary particle and not as composed of equal positive and negative charges.

of these two values of f is 1.42, so that we should expect the value of β_1 to be some 40 per cent lower for the earth than for the sun or the stars. Now it will be remembered that the mean value of β_1 for the five stars was found to be about 0.58, and for the earth 0.30. It is not excluded that some part of this observed difference of β_1 for these bodies may be due to the fact that the earth has lost most of its hydrogen while the sun and normal stars have not.

If stars exist consisting almost entirely of hydrogen, one would expect exceptionally large fields, *i. e.* 60 per cent above those of normal stars. If, as has been suggested, the very dense matter in some very dense stars consists mainly of neutrons, formed by the combination of electrons and protons, then on our hypothesis, they should possess a much smaller magnetic field than if of normal matter.

TABLE V.

Total Charge to Mass Ratio $f = \beta_1/\beta_2 = \Sigma Z/\Sigma W$.

Elements			
Element	Z	W	f
H	1	1	1.00
O	8	16	0.50
Fe	26	55.8	0.47
Pb	82	207	0.39
Compounds and Mixtures			
Substance	Composition		f
Water	H_2O		0.55
Limestone	$CaCO_3$		0.50
Olivine	$Mg_2Fe_2SiO_2$		0.48
Stellar matter	{ (35% H. by weight 65% other light elements) }		0.68

TABLE VI.

Body	k	β_1	Material	β_1/β_2	β_2
Earth	0.88	0.30	olivine	0.48	0.63
Sun (H=30G)	0.20	0.52	35% H.	0.68	0.77
5 stars	0.20	0.58	35% H.	0.68	0.85
					0.75

Since Jupiter has not lost its hydrogen, the value of β_1 , appropriate to it, will be appreciably higher than that of the earth*.

These considerations show that future experiments may be able to decide between the hypothesis that the virtual currents depend on the mass of a body or on the electric charges in it.

Using (14) we can calculate β_2 for the earth and the stars. This is most conveniently done by using the already calculated values of β_1 together with the values of $f = \beta_1/\beta_2$ from Table V. The results are given in Table VI. In all cases η is taken as unity.

* See former paper.

One sees immediately that the difference between the value of β_2 for the earth on the one hand, and the sun and the stars on the other, is much smaller than the corresponding difference between the two values of β_1 . The values are also considerably nearer unity.

If, however, Chapman's value $\eta \simeq \frac{1}{2}$, based on Schwarzschild's distribution of ω inside a star, is correct, then the value of β_2 for the sun and the stars will become about 1.5 instead of 10.75.

It is just possible that the low values of β_1 for the earth might be due to a reduction of the average external field by the induced dipoles in the core which, in Elsasser's and Bullard's theory, are the cause of the secular variation.

We conclude, therefore, that the limited and highly inaccurate data at present available, on the whole supports our second hypothesis, as expressed in equation (14).

Now the quantity $G^{\frac{1}{2}}M/e = 0.90 \times 10^{-18}$, appearing in equation (14) represents the ratio between the gravitational mass of a proton and its electrostatic charge, and is a quantity that must clearly play a fundamental part in any future unified field theory and in cosmology, just as the fine structure constant α plays an essential rôle in quantum theory. Let us write

$$\frac{G^{\frac{1}{2}}M}{e} = \epsilon. \quad (15)$$

Then, expressing i in electrostatic units, (14) becomes

$$i = -\beta_2 \epsilon \sigma_0 v, \quad (16)$$

where $v = \omega \times r$ and β_2 is experimentally determined as about 0.75 ± 0.2 .

Corresponding to this modification of (3) and (14), we must now modify (1). Our new equation, corresponding to our second hypothesis, is clearly

$$P = \frac{1}{2} \beta_2 \epsilon E, \quad (17)$$

where E can be called the electrical angular momentum of the body defined by

$$E = \int \omega \sigma_0 p^2 dV, \quad (18)$$

where p is the perpendicular distance of a volume element dV from the axis.

This general form would allow one, for instance, to calculate P for a star with any given distribution of elements of varying nuclear constitution in its interior.

The second hypothesis has also the theoretical advantage of expressing the dipole moment in terms of electric as well as mechanical properties of matter (Tzu 1947, Arley 1948).

A final highly speculative theoretical argument may be permitted. If the magnetic field of massive rotating bodies has a fundamental origin, then it seems plausible to suppose that the field is in some way bound

up with the problem of the relativity of rotational motion. Now it is already proved by Schiff's analysis of Oppenheimer's paradox (Schiff 1939), that an observer near a fixed spherical condenser, who is rotating relative to the galaxies (*i. e.* in "absolute" rotation), has to introduce fictitious electric currents, which are everywhere equal and anti-parallel to the real convective currents due to the motion of the charges of the condenser relative to himself*. These fictitious currents have the magnetic field of ordinary currents, but are not associated with the movement of real charges. Schiff's fictitious currents are in form, but not of course in magnitude, not unlike the virtual currents which we have introduced as the simplest way of explaining the observed facts of the magnetism of rotating bodies. So we see that the conception of virtual or fictitious currents is by no means foreign to the conceptions underlying the treatment of absolute rotation.

Of course, Schiff's virtual currents introduce nothing essentially new into physics and vanish for an uncharged condenser (*i. e.* for a neutral body); they are merely one aspect of conventional electromagnetic theory and so cannot provide an explanation of our supposed new phenomena expressed by (16), which must, of course, essentially be connected with gravitational phenomena. To explain this new phenomenon some new feature must be introduced. This new feature must clearly introduce some asymmetry between positive and negative electricity—to explain the actual direction of the earth's field, *i. e.* the negative sign in (1)†.

It will be remembered that in the early fundamental theories of Schuster, Sutherland, Wilson and Swann, the essential asymmetry is introduced in the form of an arbitrarily assumed difference of the order of 10^{-22} between the forces between positive and negative charges. Though these theories are quite untenable and are long since abandoned, they serve to emphasize the necessity of introducing some asymmetry between positive and negative charges.

Guided by the analogy of Schiff's theorem, we can perhaps usefully introduce this desired asymmetry by supposing that the virtual currents that have to be introduced by a rotating observer are slightly less (*i. e.* by the order of ϵ) for positive than for negative charges. So for a neutral body, *i. e.* one with equal numbers of real positive and negative charges, the postulated virtual currents due to the rotation will not cancel exactly the real convective currents, but leave a small excess negative virtual current.

* These fictitious currents come into existence by the distortion of the observer's metric by the rotation of the galaxies relative to him. In a closely similar way, a rotating observer has to introduce the fictitious Coriolis and centrifugal forces (Thirring).

† In almost all "specific" theories, this essential asymmetry enters through the differences of *mass* of the negatively charged electron and positively charged proton.

Or to express this suggestion more simply and vaguely, it seems possible that the origin of the field of a rotating body, as observed by an observer rotating with it, lies in a slight difference, of magnitude ϵ , in the behaviour of positive and negative charges when in absolute rotation.

The fact that the magnitude of the difference $\epsilon = G^{\frac{1}{2}}M/e$ is proportional to the square root of the gravitational constant, shows that the phenomenon is essentially connected with gravitation and so would vanish in the limit of $G=0$, just as quantum phenomena vanish in the limit of $\hbar=0$.

If Dirac (1937) is correct in supposing that the large non-dimensional number $e^2/G17^2 = 1/\epsilon^2$ is not a constant, but has increased linearly with the age of the universe, it follows that ϵ was much larger when the world was very young, and consequently that rotating bodies had a much larger magnetic field for a given angular momentum than today.

No attempt will be made in this paper to discuss the obvious difficulties of extending any fundamental theory of the above type to systems other than simple nearly spherical bodies, *e.g.* to planetary system, nebulae, etc.

APPENDIX

THE MAGNETIC FIELD OF WHITE DWARFS.

The writer pointed out that if White Dwarfs were formed by the collapse of main sequence stars, one would expect them to possess magnetic fields of the order of a million gauss, and that possibly the great width of the spectral lines of most of these stars might in fact be due to this cause. This arises because one would expect the angular momentum of the star and so, according to relation (1), its magnetic dipole, to be conserved during the collapse. Owing to the small final radius of the collapsed star, a very large magnetic field at the surface would be expected.

In a private communication, Babcock has reported observations of the Balmer lines of 40 Eridani B, using an analyser for circularly polarized light. He found no sign of any Zeeman effect. A. D. Thackeray (1947) has obtained spectra of Wolf 1346, and has also failed to find any evidence for any Zeeman effect.

It is of course possible that White Dwarfs are not formed by the collapse of normal stars. This is the view of Schatzman (1947), who considers that they have probably a quite different cosmological origin. Alternatively perhaps some mechanism such as planet or ring formation may have operated to remove a large part of the angular momentum. It seems, however, rather unlikely that such a mechanism can always come into play when a star collapses. It seems widely accepted that novæ and supernovæ are due to such a process of collapse, brought about perhaps by exhaustion of hydrogen, or possibly by the setting in of energy loss by neutrino emission, as in the theory of Gamow and Schoenberg. If such processes do exist the resulting small and dense stars should have a large magnetic field if they were originally rotating as fast, say, as

the sun. It is of course possible that stars of this kind do exist, but then they are not White Dwarfs. An alternative possibility is that the matter in the interior of these stars consists mainly of neutrons and that consequently (according to (14)), the magnetic field is small.

I wish to express my thanks to many of my colleagues, in particular to Professor Rosenfeld, Professor Freundlich, Professor Chapman, Mr. Runcorn and Mr. Tzu, for valuable discussions of many aspects of this work.

REFERENCES.

- ALFVEN, 1947, *Phys. Rev.*, **72**, 88.
 ARLEY, 1948, *Nature, Lond.*, **161**, 596.
 BABCOCK, H. W., 1947a, *Astrophys. J.*, **105**, 105.
 BABCOCK, H. W., 1947b, *Publ. Astr. Soc. Pacif.*, **59**, 112.
 BABCOCK, H. W., 1947c, *Phys. Rev.*, **72**, 83.
 BABCOCK, H. W., 1948a, *Phys. Rev.*, **74**, 489.
 BABCOCK, H. D., 1948b, *Publ. Astr. Soc. Pacif.*, **60**, 244.
 BECKER, 1942, *Sterne und Sternsysteme*.
 BLACKETT, 1947, *Nature, Lond.*, **169**, 658.
 BULLARD, 1948, *Mon. Not. R. Astr. Soc., Geophys. Suppl.*, **5**, 248; also short report in *The Observatory*, **68**, 144.
 CHAPMAN, 1948a, *Proc. Phys. Soc.*, **61**, 95.
 CHAPMAN, 1948b, *Mon. Not. R. Astr. Soc.* **108**, 236.
 CHAPMAN, 1948c, *Nature, Lond.*, **161**, 52.
 CHAPMAN, 1948d, *Annales de Geophysique*, **4**, 109.
 COWLING, 1933, *Mon. Not. R. Astr. Soc.*, **94**, 40.
 COWLING, 1945, *Mon. Not. R. Astr. Soc.*, **105**, 166.
 DEUTSCH, 1947, *Astrophys. J.*, **105**, 283.
 DIRAC, 1937, *Nature, Lond.*, **139**, 323.
 ELSASSER, 1941, *Phys. Rev.*, **60**, 159.
 ELSASSER, 1946, *Phys. Rev.*, **69**, 106; *Ibid.*, **70**, 212.
 HALES, and GOUGH, 1947, *Nature, Lond.*, **160**, 746.
 RUNCORN, 1948, *Discussion at R. Astr. Soc.* 27 Feb. reported in *Nature, Lond.*, **161**, 462; *Proc. Phys. Soc.* **61**, 373.
 SCHATZMAN, 1947, *Annales d'Astrophys.*, **10**, 93.
 SCHIFF, 1939, *Proc. Nat. Acad. Sci.*, **25**, 391.
 SCHWARZSCHILD, 1947, 1947, *Astrophys. J.*, **106**, 427.
 THACKERAY, 1947, *Mon. Not. R. Astr. Soc.*, **107**, 463.
 THIESSEN, 1946, *Annales d'Astrophys.*, **9**, 101.
 TZU, 1947, *Nature, Lond.*, **160**, 746.
 WESTGATE, 1933, *Astrophys. J.*, **78**, 46; 1934, **79**, 357.
 WILSON, 1923, *Proc. Roy. Soc. A*, **104**, 451.

XII. Canonical Transformations of the Hamiltonian in Meson Field Theory.

By K. J. LE COUTEUR and L. ROSENFELD,
University of Manchester *.

[Received October 15, 1948.]

IN a recent note, Dyson (1948) pointed out that by means of a certain canonical transformation it was possible to put the Hamiltonian of pseudoscalar meson field theory, and especially the coupling energy of this field with nucleons or leptons, into a simpler form, well-suited to bring out peculiar features of the theories of nuclear interaction and β -decay in this case, which would otherwise appear to be due to accidental reductions. Since Dyson himself does not go into any detail concerning these consequences of his transformation, and the wording of a remark at the beginning of his note might be understood to imply criticism of previous results, it is perhaps not superfluous to discuss these points a little more fully.

For the scalar meson theory, Dyson gives another canonical transformation, which again simplifies the corresponding expression for the Hamiltonian. We shall also discuss this case in the following; it will be seen that, in contrast to the pseudoscalar case, Dyson's transformation for the scalar case is closely related to that introduced by Stueckelberg, and especially by Møller and Rosenfeld (1940), in order to separate the static part of the nuclear interactions. An ambiguity remaining in the latter transformation will be shown to have no physical meaning; as a result, the non-static interaction of the first order, for which a certain expression had been given by Rosenfeld (1945), can be made to cancel. By a similar argument, the analogous expression for the first order interaction in the pseudovector case can be considerably reduced.

Pseudoscalar Meson Theory.

We consider the charge symmetric theory. Then in the notation of Rosenfeld (1945) the Hamiltonian for a system of nucleons, leptons and mesons in interaction is

$$H = H_n + H_l + H_p + H_{int}^{(1)} + H_{int}^{(2)} \quad . \quad . \quad . \quad . \quad (1)$$

with

$$H_n = \sum_i [\vec{\alpha}^i p^i + \rho_3^i \{M + \frac{1}{2}(M_n - M_p)\tau_3^i\}], \quad . \quad . \quad . \quad . \quad (2a)$$

* Communicated by the Authors.

$$H_\psi = \frac{1}{2} \int \{ \boldsymbol{\varphi}^2 + (\text{grad } \boldsymbol{\psi})^2 + \kappa^2 \boldsymbol{\psi}^2 \} dv, \quad \dots \quad (2b)$$

$$H_{\text{int}}^{(1)} = - \int \{ f_1 \mathbf{r} \boldsymbol{\psi} + (f_2/\kappa) \{ \mathbf{q} \boldsymbol{\varphi} + \mathbf{s} \text{grad } \boldsymbol{\psi} \} + \tilde{f}_1 \tilde{\mathbf{r}} \boldsymbol{\psi} + (\tilde{f}_2/\kappa) \{ \tilde{\mathbf{q}} \tilde{\boldsymbol{\varphi}} + \tilde{\mathbf{s}} \text{grad } \tilde{\boldsymbol{\psi}} \} \} dv, \quad \dots \quad (2c)$$

$$H_{\text{int}}^{(2)} = (1/2\kappa^2) \int (f_2^2 \mathbf{q} + \tilde{f}_2^2 \tilde{\mathbf{q}})^2 dv. \quad \dots \quad (2d)$$

The Hamiltonian H_l for the leptons is similar to H_n with $\tilde{\alpha}$, \tilde{p} in place of α , p and so on: quantities referring to the light particles are always distinguished by the superscript \sim . We write $M = \frac{1}{2}(M_p + M_n)$ and $\tilde{M} = \frac{1}{2}m$ where M_p , M_n , m , 0 are the masses of proton neutron, electron and neutrino. $\boldsymbol{\varphi}$ is the canonical conjugate of $\boldsymbol{\psi}$ satisfying the commutation rule

$$(\psi_m(x), \phi_n(x')) = i\hbar \delta(\vec{x} - \vec{x}') \delta_{mn}, \quad \dots \quad (3)$$

and so

$$d\boldsymbol{\psi}/dt = \boldsymbol{\varphi} - (1/\kappa)(f_2 \mathbf{q} + \tilde{f}_2 \tilde{\mathbf{q}}).$$

The term $H_{\text{int}}^{(2)}$ arises from the substitution of this expression for $d\boldsymbol{\psi}/dt$ in the calculation of H from the Lagrangian.

$H_{\text{int}}^{(2)}$ is not uniquely defined, for we can always add to it an invariant term $\eta \int (\mathbf{q} \tilde{\mathbf{q}} - \mathbf{s} \tilde{\mathbf{s}}) dv$, with arbitrary η , which represents a direct Fermi coupling between nucleons and leptons.

The densities \mathbf{r} , \mathbf{q} , $\tilde{\mathbf{s}}$ are defined as

$$\left. \begin{aligned} \mathbf{r} &= \sum_i \boldsymbol{\tau}^i \rho_2^i \delta(\vec{x} - \vec{x}^i), \\ \mathbf{q} &= \sum_i \boldsymbol{\tau}^i \rho_1^i \delta(\vec{x} - \vec{x}^i), \\ \tilde{\mathbf{s}} &= \sum_i \boldsymbol{\tau}^i \sigma^i \delta(\vec{x} - \vec{x}^i). \end{aligned} \right\} \dots \quad (4)$$

Following Dyson, we now transform H to the form

$$H' = e^{iS/\hbar} H e^{-iS/\hbar} \quad \dots \quad (5)$$

with

$$S = -(1/\kappa) \int (f_2 \mathbf{q} + \tilde{f}_2 \tilde{\mathbf{q}}) \boldsymbol{\psi} dv. \quad \dots \quad (6)$$

$H_{\text{int}}^{(2)}$ is invariant under this transformation but is cancelled by terms arising from the transformation of H_ψ and $H_{\text{int}}^{(1)}$. The new Hamiltonian is, correct up to terms of second order in the coupling coefficients f ,

$$\begin{aligned} H &= H_n + H_l + H_\psi \\ &- \left(f_1 + 2f_2 \frac{M}{\mu} \right) \int \mathbf{r} \boldsymbol{\psi} dv - \left(\tilde{f}_1 + 2\tilde{f}_2 \frac{\tilde{M}}{\mu} \right) \int \tilde{\mathbf{r}} \boldsymbol{\psi} dv \\ &- f_2 \frac{M_n - M_p}{\mu} \sum_i \rho_2^i \rho_3^i - \tilde{f}_2 \frac{m}{\mu} \sum_i \tilde{\rho}_2^i \tilde{\rho}_3^i, \quad \dots \quad (7) \end{aligned}$$

where μ is the rest mass of the meson. Disregarding the small terms of the last line of (7), we thus see that Dyson's transformation has the effect of removing from the original Hamiltonian all coupling terms depending on

the constants f_2, \tilde{f}_2 . In the first place, we see that apart from small non-static effects proportional to $f_2(M_n - M_p)/\mu$, the nuclear forces depend on the parameter

$$\left(f_1 + 2f_2 \frac{M}{\mu}\right)^2 = \left(\frac{2M}{\mu}\right)^2 \left\{f_2^2 + f_2 f_1 \frac{\mu}{M} + \left(f_1 \frac{\mu}{2M}\right)^2\right\}. \quad (8)$$

Now, the nuclear potential was calculated by Rosenfeld (1945) as the sum of two terms classified as

(i) static forces,

(ii) non-static forces of first order in the velocity of the nucleons,

and terms classified as of higher order in the velocity were omitted as of dubious physical significance. The terms (i), (ii) correspond to the first two terms of the expansion (8) and the omitted terms presumably include the third term of (8). There is therefore no discrepancy between the results obtained by the two different methods; the similarity of form found by direct computation for the static and first order terms appears as an immediate consequence of Dyson's transformation.

Likewise, it follows from (7) that the lifetime t_0 of the pseudoscalar meson at rest is determined by the parameter $(\tilde{f}_1 + \tilde{f}_2 m/\mu)$, actually

$$\frac{1}{t_0} = c\kappa \left\{ \frac{\tilde{f}_1}{\sqrt{\hbar}} + \frac{\tilde{f}_2}{\sqrt{\hbar}} \frac{m}{\mu} \right\}^2 \quad (9)$$

in agreement with Sakata's (1941) and Chang's (1942) direct calculation*.

As regards the theory of β -decay, the situation is obscured by an error in Sakata's (1941) direct treatment. From the transformed Hamiltonian (7) it is immediately apparent that the decay constant will involve the factor

$$\left\{f_1 + f_2 \frac{M_p + M_n}{\mu}\right\}^2 \left\{\tilde{f}_1 + \tilde{f}_2 \frac{m}{\mu}\right\}^2 \quad (10)$$

This agrees, of course, with the conclusion arrived at by Nelson (1941) using an argument essentially equivalent with Dyson's transformation. The apparent discrepancy with Sakata's result can be traced to an inconsistency in the latter's calculations: the error lies in his replacement of the interaction Hamiltonian (70 d) + (78 d) by the approximation (79 d). It is better to simplify (78 d) by means of the substitutions (our \vec{s} and \vec{r} correspond to Sakata's \vec{M} and \vec{W})

$$\text{div } \vec{s} = -\frac{2Mc}{\hbar} \vec{r}, \quad \text{div } \vec{s} = -\frac{2\tilde{M}c}{\hbar} \vec{r}, \quad (11)$$

which, as discussed by Rosenfeld (1945), are valid approximations in

* The corresponding formula (89 d) in Sakata's (1941) paper has a trivial error of sign. In Rozental's (1941) slightly different notation, however, the sign of the term in \tilde{f}_2 must be changed.

problems where the nucleons and leptons may be treated as free. Then we obtain in place of Sakata's (79 d),

$$H = - \int dv \, \bar{v} \left(f_1 + 2f_2 \frac{M}{\hbar} \right) \left(\tilde{f}_1 + 2\tilde{f}_2 \frac{\tilde{M}}{\mu} \right) \mathbf{r} \tilde{\mathbf{r}}, \quad . \quad . \quad . \quad (12)$$

which leads to the same results as our Hamiltonian (7).

It must be stressed, however, that owing to the essential ambiguity of the original Hamiltonian (2 d), it would always be possible, by a suitable addition of direct couplings between nucleons and leptons, to introduce into the expression for the decay constant another combination of constants than that derived above. This possibility cannot be dismissed just by saying that it is tantamount to abandoning the idea of the meson field as an intermediary in the mechanism of β -decay, since it is this mechanism itself that gives rise to the ambiguity in the form of the interaction term (2 d). However, since it has been discovered that the meson responsible for the nuclear forces and (possibly) for the β -decay have a much shorter life time than was thought formerly, the whole issue has lost any physical significance.

Scalar Meson Theory.

We now have source densities

$$\left. \begin{aligned} \mathbf{l} &= \Sigma^i \boldsymbol{\tau}^i \rho_3^i \delta(\vec{x} - \vec{x}^i), \\ \mathbf{n} &= \Sigma^i \boldsymbol{\tau}^i \delta(\vec{x} - \vec{x}^i), \\ \vec{\mathbf{m}} &= \Sigma^i \boldsymbol{\tau}^i \rho_1^i \sigma^i \delta(\vec{x} - \vec{x}^i), \end{aligned} \right\} . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

and the interaction Hamiltonian becomes, in place of (2 c), (2 d)

$$\begin{aligned} H_{\text{int}}^{(1)} &= -f_1 \int \mathbf{l} \boldsymbol{\psi} \, dv - \frac{f_2}{\kappa} \int \{ \mathbf{n} \boldsymbol{\phi} + (\vec{\mathbf{m}} \text{ grad } \boldsymbol{\psi}) \} \, dv, \\ &- \tilde{f}_1 \int \tilde{\mathbf{l}} \boldsymbol{\psi} \, dv - \frac{\tilde{f}_2}{\kappa} \int \{ \tilde{\mathbf{n}} \boldsymbol{\phi} + (\vec{\tilde{\mathbf{m}}} \text{ grad } \boldsymbol{\psi}) \} \, dv, \quad . \quad . \quad (14 c) \end{aligned}$$

$$H_{\text{int}}^{(2)} = \frac{1}{2\kappa^2} \int (f_2 \mathbf{n} + \tilde{f}_2 \tilde{\mathbf{n}}) \, dv. \quad . \quad . \quad . \quad . \quad . \quad . \quad (14 d)$$

Following Dyson, we now apply a transformation of the form (5) with

$$S = - \frac{1}{\kappa} \int (f_2 \mathbf{n} + \tilde{f}_2 \tilde{\mathbf{n}}) \boldsymbol{\psi} \, dv. \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

The new Hamiltonian is, correct up to terms of second order in the coupling coefficients,

$$\begin{aligned} H &= H_n + H_l + H_p \\ &- f_1 \int \mathbf{l} \boldsymbol{\psi} \, dv - \tilde{f}_1 \int \tilde{\mathbf{l}} \boldsymbol{\psi} \, dv \\ &- f_2 \frac{M_n - M_p}{\mu} \Sigma \rho_3^i (\psi_1 \tau_2 - \psi_2 \tau_1)^i - \tilde{f}_2 \frac{m}{\mu} \Sigma \tilde{\rho}_3^i (\tilde{\psi}_1 \tilde{\tau}_2 - \tilde{\psi}_2 \tilde{\tau}_1)^i. \quad . \quad (16) \end{aligned}$$

Again, the main terms in f_2, \tilde{f}_2 are removed from the Hamiltonian; but the last two terms are quite different from the corresponding terms of (7), the reason being that ρ_3 commutes with \mathbf{n} but anticommutes with \mathbf{q} . To interpret these terms we introduce the complex wave-function $\psi = \psi_1 + i\psi_2$ and the operator $\Pi = \frac{1}{2}(\tau_1 - i\tau_2)$.

Then

$$\tau_2\psi_1 - \tau_1\psi_2 = -i\Pi^+\psi^+ + i\Pi\psi,$$

while

$$\tau_1\psi_1 + \tau_2\psi_2 = \Pi^+\psi^+ + \Pi\psi.$$

Then

$$H = H_n + H_l + H_p$$

$$\begin{aligned} & -\Sigma\rho_3^i\left\{\left(f_1 - if_2\frac{M_n - M_p}{\mu}\right)\Pi^+\psi^+ + \left(f_1 + if_2\frac{M_n - M_p}{\mu}\right)\Pi\psi + f_1\tau_3\psi_3\right\} \\ & -\Sigma\rho_3^i\left\{\left(\tilde{f}_1 - i\tilde{f}_2\frac{m}{\mu}\right)\tilde{\Pi}^+\psi^+ + \left(\tilde{f}_1 + i\tilde{f}_2\frac{m}{\mu}\right)\tilde{\Pi}\psi + \tilde{f}_1\tau_3\psi_3\right\}. \quad (17) \end{aligned}$$

It follows that the lifetime of a scalar meson at rest is determined by

$$\left|\tilde{f}_1 + i\tilde{f}_2\frac{m}{\mu}\right|^2 = \tilde{f}_1^2 + \tilde{f}_2^2\left(\frac{m}{\mu}\right)^2, \quad (18)$$

in agreement with Sakata's result. Similarly the β -decay of the nucleons depends on the product

$$\left\{f_1^2 + f_2^2\left(\frac{M_n - M_p}{\mu}\right)^2\right\}\left\{\tilde{f}_1^2 + \tilde{f}_2^2\left(\frac{m}{\mu}\right)^2\right\}. \quad (19)$$

Although at first sight one would expect Dyson's transformation again to annihilate all f_2 effects we see that, in fact, there remain some small terms in f_2^2 and \tilde{f}_2^2 , independent of the nucleon and lepton velocities.

Passing now to the discussion of the nuclear forces, it will be sufficient and more convenient, for our present purpose, to treat the case of the neutral theory, with interaction Hamiltonian

$$H_{\text{int}} = -f_1 \int \bar{\psi} \psi dv - \frac{f_2}{\kappa} \int \{n\phi + \vec{m} \text{ grad } \psi\} dv + \frac{f_2^2}{2\kappa} \int n^2 dv. \quad (20)$$

Let us first elucidate the relation between Dyson's transformation and that used by Møller and Rosenfeld (1940) to separate off the static part ψ_0, ϕ_0 of the meson field,

$$\phi_0 = \frac{f_2}{\kappa} n, \quad \psi_0 = f_1 \Sigma^i \rho_3^i \phi(\vec{x} - \vec{x}^i). \quad (21)$$

In this notation, the operator of Dyson's transformation, according to (15), may be written

$$S_2 = -\int \phi_0 \psi dv, \quad (22)$$

whereas that used by Møller and Rosenfeld is defined by

$$S = S_1 + S_2 \quad (23)$$

with $S_1 = \int \psi_0 \phi \, dv$ (24)

Since, however, S_1 and S_2 do not commute, the latter transformation is not uniquely determined. Instead of

$$H^a = e^{iS_1/\hbar} H e^{-iS_1/\hbar}, \quad (25 a)$$

the new Hamiltonian

$$H^b = e^{iS_1/\hbar} e^{iS_2/\hbar} H e^{-iS_2/\hbar} e^{-iS_1/\hbar}, \quad (25 b)$$

for instance, would do equally well; and it is this last choice that corresponds to Dyson's method, since it amounts to first applying Dyson's transformation (22). By expanding the exponentials we readily find that (25 a) and (25 b), up to the second order, only differ in the second order contributions from the nucleon kinetic energy H_n , viz.,

$$\frac{1}{2} \left(\frac{i}{\hbar} S_1, \left(\frac{i}{\hbar} S_2, H_n \right) \right) + \frac{1}{2} \left(\frac{i}{\hbar} S_2, \left(\frac{i}{\hbar} S_1, H_n \right) \right) \quad . . . (26 a)$$

and

$$\left(\frac{i}{\hbar} S_1, \left(\frac{i}{\hbar} S_2, H_n \right) \right), \quad (26 b)$$

respectively. The difference between (26 a) and (26 b) may be written

$$\frac{1}{2} \left(\left(\frac{i}{\hbar} S_1, \frac{i}{\hbar} S_2 \right), H_n \right) = \frac{i}{2\hbar} \left(\int \psi_0 \phi_0 \, dv, H_n \right). \quad (27)$$

Thus the apparent ambiguity in the second order terms (26) is of no physical significance, for the different expressions (26 a, b) can be transformed into each other by use of another transformation of operator $\frac{1}{2} \int \psi_0 \phi_0 \, dv$ which is equivalent to a trivial redefinition of the dynamical variables describing the free nucleons.

Now, Dyson's method in the scalar case immediately shows that if we neglect terms depending on $(M_n - M_p)/\mu$ the nuclear potential only depends on the constant f_1 , and therefore reduces to the static part

$$-\frac{1}{2} \sum_{i, k} f_1^2 \phi^{(ik)}.$$

The fact that in applying the transformation (25 a) Rosenfeld (1945) found an additional contribution of the first order in the nucleon velocities is not in contradiction with this conclusion: this additional term is just * the expression (27); as explained above, it has no physical meaning and may simply be cancelled.

A similar situation arises in the pseudovector meson theory. It can be seen that in this case also the transformation corresponding to (25 b), with

$$S_1 = \int \vec{F}^0 \vec{U} \, dv, \quad S_2 = - \int \vec{U}^0 \vec{F} \, dv,$$

while being equivalent (to the second order of approximation) with (25 a), leads to a simpler result: the meaningless difference analogous to (27)

* In Rosenfeld's paper, explicit mention should have been made of the fact that the terms corresponding to (26 a) vanish in both scalar and pseudovector theory.

turns out to be the last term in the expression for the non-static interaction given by Rosenfeld (1945), equation (24). This term may accordingly be cancelled. In the vector and pseudoscalar cases, on the other hand, it has been pointed out by Møller and Rosenfeld (1940), § 4, that the expression corresponding to (27) vanishes, so that there does not arise any ambiguity in those cases.

REFERENCES.

- CHANG, T. S., 1942, *D. Kgl. Danske Vid. Selsk.*, **19**, No. 10.
DYSON, F. J., 1948, *Phys. Rev.*, **72**, 929.
MØLLER, C., and ROSENFELD, L., 1940, *D. Kgl. Danske Vid. Selsk.*, **17**, No. 8.
NELSON, E. C., 1942, *Phys. Rev.*, **60**, 830.
ROSENFELD, L., 1945, *D. Kgl. Danske Vid. Selsk.*, **32**, No. 13.
ROZENTAL, S., 1941, *D. Kgl. Danske Vid. Selsk.*, **18**, No. 7.
SAKATA, S., 1941, *Proc. Phys. Math. Soc., Japan*, **32**, 291.

XIII. *Anomalous Dielectric Properties of Polycrystalline Titanates of the Perovskite Type.*

By J. R. PARTINGTON, G. V. PLANER, and I. I. BOSWELL *.

[Received August 11, 1948.]

Barium titanate and its solid solutions with a number of related compounds have attracted considerable interest since the discovery, some few years ago, of their exceptionally high permittivities. Substances of this class, possessing structures of the perovskite type, are ferroelectric at temperatures up to the Curie point; at the latter, a second order transition from a distorted to a cubic crystal structure takes place. The permittivity rises to a maximum in the temperature region of the transition, attaining in some cases values as high as 10,000.

The structure of polycrystalline barium titanate is tetragonal below and cubic above the temperature of the Curie point, the transition taking place in the region of 120° C. (Megaw 1946). At approximately 10° C. a second, lower maximum is observed in the permittivity-temperature curve, the cause of which is as yet unexplained. In solid solutions of barium and strontium titanates, Curie points may be obtained over a wide range of temperatures, depending on the relative proportions of the constituents. The Curie point of a mixture of 75 per cent by weight of barium titanate and 25 per cent of strontium titanate, for example, lies between 10° C. and 25° C. As in the case of the barium compound, the structure of barium strontium titanate is cubic in the high temperature region, and on cooling below the Curie point, it breaks up into tetragonal domains. Within a region close to the temperature corresponding to the permittivity maximum, the two modifications have been shown to

* Communicated by G. V. Planer.

co-exist. It has been suggested that this effect may be due to local anisotropic stresses on the crystallites in the material, resulting in a distribution of transition points over a narrow range of temperatures (Megaw 1947).

By analogy with the ferromagnetic case, the permittivity of ferroelectric materials, measured with a weak alternating field during the superposition of a strong unidirectional electric field, is known as the "reversible permittivity." This is a function of the constant polarization of the dielectric, and is thus a measure of the ease with which further polarization can be effected in the material. The "reversible permittivity" falls with increasing strength of the superimposed field and rises to a maximum as the latter is reduced to zero.

It has been reported by several investigators (Wul 1946, Roberts 1947, Reddish, Plessner and Willis Jackson 1946) that the A.C. permittivity of barium and barium strontium titanate is, in fact, reduced by exposure to strong unidirectional fields, at temperatures below as well as above the Curie points. There appeared, however, to be no information on the variation with time of the permittivity during the application of a superimposed field and shortly after its removal, and such an investigation was considered of interest for the following reasons. It was anticipated that a modification in the permittivity might be detectable after the exposure to strong D.C. fields, because of the possibility of a remanent polarization with a sufficiently prolonged period of decay. Residual effects of this type were considered possible in view of the slow discharge rates reported in the case of condensers containing Rochelle salt, another ferroelectric substance, as the dielectric (Valasek 1922, Cady 1946).

The comparative instability of the crystal structure in the region of the transition from the tetragonal to the cubic form was regarded as another possible cause for a modification in the permittivity by electric stresses.

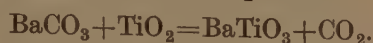
At temperatures below the Curie point, the relation between the polarization and the applied electric field is represented by a hysteresis loop, while above, it becomes linear. The ferroelectric influence continues, however, to make itself felt within a temperature region extending well above the Curie point, and for this reason it was considered of interest to study the residual effects above as well as below the transition temperature.

EXPERIMENTAL.

The experimental method depended in principle on the determination of variations of the permittivity of titanates with time, during as well as after the application of strong unidirectional electric fields. For this purpose, the curves relating the permittivity to the temperature were at first determined for various compositions. Measurements of the permittivity changes due to the high tension treatments were then carried out at different temperatures below and above the Curie points. The changes in permittivity immediately after the short application of a strong field were first studied; later the variations with time of the permittivity during and after longer applications of the tension were examined.

Preparation of samples.

The titanate specimens were prepared by the reaction between titanium dioxide and an alkaline earth carbonate. In the case of barium titanate the reaction proceeds according to the equation



The requisite quantities of the powders were ground in a porcelain mill for several hours, together in some cases with small quantities of a binder, such as bentonite, and other additives. The mixture was pre-fired at 1250°C. and the powder then pressed into flat disks in a steel die, at a pressure of 8000 lbs./sq. in. The plates were generally fired at 1400°C. in an electric furnace, and in some cases at temperatures between 1350°C. and 1400°C. Silver electrodes were applied to the fired disks by painting a suspension of silver oxide in acetone on to the faces and heating the specimens at 700°C. Electrical contact was made to the electrodes by wires soldered on to the silver coatings. The samples were rinsed with methylated spirit and dried at 100°C., in some cases for several days. Shortly before taking measurements the disks were transferred to a desiccator containing calcium chloride and allowed to cool to room temperature. A protective coating was applied, in some cases, by dipping the samples into molten wax while hot, after drying in an oven for several days.

Specimens of the following compositions were prepared in this way :

- (1) Barium titanate.
- (2) Barium strontium titanate ($\text{BaTiO}_3 : \text{SrTiO}_3 = 75 : 25$, by weight).
- (3) Barium titanate with different additives.
The following substances were incorporated in the mixes :
 - (a) Excess titanium dioxide (3 per cent with respect to the weight of barium titanate),
 - (b) cerium dioxide (various proportions),
 - (c) cerium dioxide and excess barium carbonate,
 - (d) aluminium trioxide (2.5 per cent).
 - (e) vanadium pentoxide (2.5 per cent).

Electrical measurements were carried out on the above specimens, as well as on a number of samples kindly supplied by Messrs. United Insulator Co., Ltd., and Taylor Tunnicliff & Co., Ltd.

Electrical Measurements.

The curves relating the permittivity and loss tangents of the specimens to the temperature were first of all determined. The samples were then maintained at various constant temperatures, the permittivity and dielectric losses being determined before and at various time intervals after the application of a high tension. Accurate temperature control was essential during these measurements by reason of the high temperature

coefficient of permittivity of the materials. The thermostat generally used comprised a small metal chamber containing the sample, through which dried air was passed. This was derived from two separate streams, one of which was passed over an electric heating element, the other through a coil immersed in a cold water bath. In the case of measurements below room temperature the cold air stream was led through a bath containing ice. The temperature of the final air stream could be regulated accurately by means of a mixing valve controlling the proportions of hot and cold air entering the thermostat chamber. In this way the temperature could be controlled well within $\pm 0.05^\circ \text{C.}$ for temperatures from 20 to 50°C. , and within $\pm 0.1^\circ \text{C.}$ for temperatures between 50°C and 120°C. The temperature was read on a thermometer partly inside the thermostat chamber; checks were made in several cases with a thermocouple soldered onto one of the silver electrodes of the sample.

In the case of measurements at temperatures above 120°C. a thermostatically controlled electric oven was used.

Before readings were recorded, the samples were in all cases allowed to attain constant temperature in the thermostat during 5 to 60 minutes, until approximate constancy in the permittivity values was obtained.

A bridge method was used for the majority of the permittivity determinations, the frequency being generally 1 kc./s., and the measuring tensions 1 to 5V./3 mm. Variations in the capacity of $0.2 \mu\text{F.}$ could readily be detected by this method, the capacities of the test specimens ranging from 1,000 to 10,000 $\mu\text{F.}$

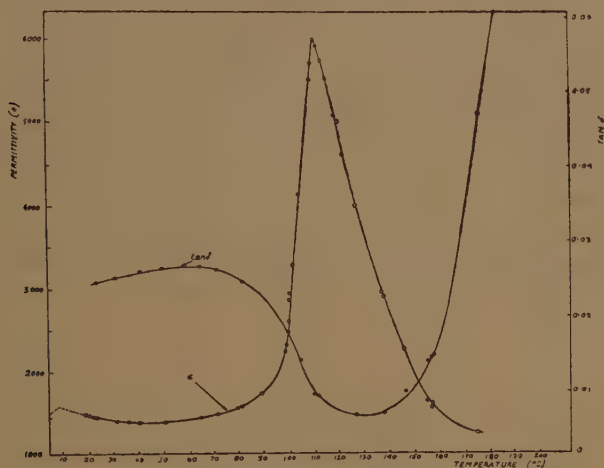
The high tension, supplied by a power pack, was measured with an electrostatic voltmeter. After exposure to the field, the electrodes of the test specimens were generally connected to earth for some seconds.

The material first examined was barium titanate (samples A.1, A.2 and A.3). Fig. 1 shows the curves relating the permittivity and loss tangent to the temperature. The Curie point for sample A.1 is seen to be at approx. 111°C. The permittivity values measured are somewhat lower than the true permittivity of barium titanate, owing to the porosity of the polycrystalline specimens (Reddish, Plessner and Willis Jackson 1946, Rushman and Strivens 1946, 1947). Since the absolute values of the permittivity were not required for the present purpose, the measurements were not corrected for this effect. The porosity was, moreover, reduced as far as possible by the use of high firing temperatures in the preparation of the specimens.

Fig. 2 shows typical curves obtained for the variations with time for the permittivity and loss tangent after the short application of a high tension. At temperatures below the Curie point the permittivity was found to have an increased value after removal of the field and to revert, at first rapidly, later more slowly, to its original value. The decay extended over periods of up to 50 minutes from the time of cessation of the tension. The effects became more pronounced for longer exposures to the field, and on increasing the temperature, being greatest just below the Curie point.

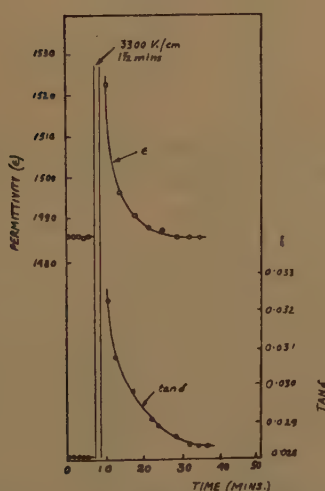
A marked increase in permittivity was also observed at temperatures corresponding to the portion of the plateau in the permittivity-temperature curve, in which the temperature coefficient is slightly negative.

Fig. 1.



Sample A.1 [BaTiO₃].

Fig. 2.



Sample A.1 [BaTiO₃]. D.C. Field 3300 V./cm., applied for 1½ mins.
Temperature 24.5° C.

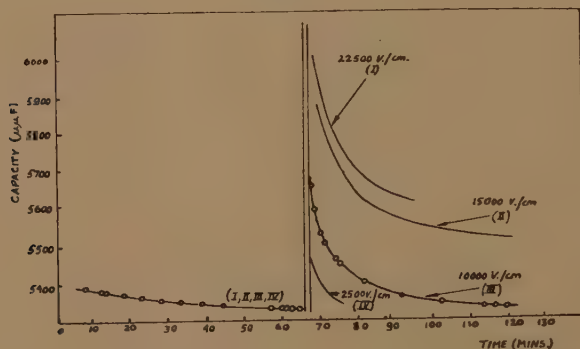
Analogous results were obtained when the electrodes of the test specimens were earthed for some seconds or for two minutes respectively, after the application of the field.

The loss tangent varied in each case in a similar manner to the permittivity, approaching very nearly its original value after periods of up to 50 minutes.

The value of the permittivity and loss tangent prior to the application of the field in fig. 2 differs somewhat from the corresponding value in fig. 1; these deviations are due to the "cyclic drift" in titanates, *i. e.* the small change observed in the electrical properties owing to temperature variations.

The effect of the magnitude of the D.C. field was next examined. Fig. 3 shows the permittivity-time curves obtained on the application of tensions of 2500, 10,000, 15,000 and 22,500 V./cm., in the case of sample A.2. The period of application was in all cases $1\frac{1}{2}$ minutes. To allow for slight differences in the capacity prior to the voltage treatment due to cyclic drift, curves (1), (2) and (4) are displaced along the ordinate, so that the initial portions of the curves coincide. The effects are seen to become larger with increasing field strengths, and the maximum change in capacity observed at 22,500 V./cm. was as high as 13 per cent of the initial value.

Fig. 3.

Sample A.2 [BaTiO₃]. Temperature 28° C.

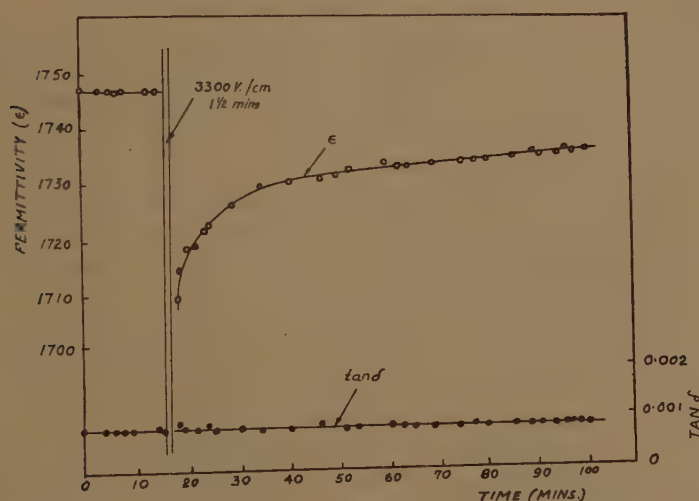
At temperatures above the Curie point, a temporary decrease in permittivity was observed after the application of a high tension, and this effect will be discussed below

To ascertain that no heating of the specimen took place on application of the high tension, the temperature was measured with a thermocouple soldered onto one of the silver electrodes, the sample being contained in the air thermostat. No variation in temperature was observed during or after a 10 minutes application of a D.C. field of 3300 V./cm. The resistance values of sample A.1. were further measured at various temperatures at a tension of 3300 V./cm. These were seen to range from a value above 3×10^4 Megohms at room temperature to 4×10^2 Megohms at 100° C. The conduction current was therefore negligible and could not produce any noticeable heating effects.

The material next examined was a solid solution of barium and strontium titanate (BaTiO₃:SrTiO₃ 75:25, by weight; samples B.1 and B.2). The curves relating the permittivity and loss tangent to the temperature were again determined and the Curie point was 21° C. The variations

in permittivity due to short applications of a D.C. field were now examined. At temperatures below the Curie point the permittivity-time curves were similar to those obtained for barium titanate below the transition temperature, a temporary increase in permittivity being observed. Above the Curie point the permittivity had a decreased value on cessation of the tension, and reverted at a diminishing rate to its initial magnitude. Fig. 4 shows a typical curve obtained in this region. The values of the loss tangents, which were low at temperatures well above the Curie point, remained effectively constant after voltage treatments; closer to the transition temperature the change in the permittivity was again accompanied by a temporary increase in $\tan \delta$.

Fig. 4.



Sample B.1 [(BaSr)TiO₃]. D.C. Field 3300 V./cm., applied for 1½ mins.
Temperature 70° C.

Similar curves were obtained at various temperatures between 25° C. and 80° C., and the decreases in permittivity were seen from these to become greater at temperatures approaching the Curie point. The effects were again found to be largest in the case of higher tensions and prolonged periods of application. Within two to three degrees above the temperature of the permittivity maximum (*i. e.* within the region of co-existence of the cubic and tetragonal structures), a temporary increase in permittivity was, however, observed, as in the tetragonal range.

The samples next examined were those consisting of barium titanate incorporating small quantities of various additives. The addition of cerium dioxide (Preparation 3b and c) resulted in a displacement of the Curie point to lower temperatures, depending on the percentage of the oxide used, and this may be ascribed to the formation of solid solutions of barium titanate and barium cerate. Measurements of the variation in permittivity due to the application of a D.C. field were taken at

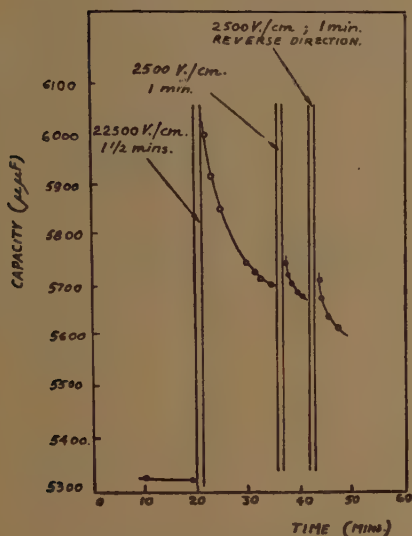
temperatures below and above the Curie point of a specimen with a transition temperature of 50°C . In the former case a temporary increase was again observed, while in the latter a decrease took place, the curves obtained being similar to those shown in figs. 2 and 4, respectively. On the application of higher tensions to specimens of this type a permanent modification in the permittivity was observed under certain conditions, which will be discussed later. No modifications in the Curie point were noted in the samples containing alumina, vanadium oxide or excess titania. The variations in the permittivity due to the application of fields of up to 5000 V./cm. were in all cases of the same type as the temporary changes observed in the previous samples. Similar results were also obtained in the case of a solid solution of barium titanate and a zirconate, exposure to the field resulting again in an increase in permittivity below, and a decrease above the Curie point.

The effect of varying the frequency and magnitude of the A.C. measuring field was next examined. Permittivity measurements were taken before and after the application of high tensions, using test frequencies of 500 c./s. , 15 kc./s. , 130 kc./s. and 1 Mc./s. , and voltages of 5 V./3mm. and 1 V./3 mm. A resonance method was used for the measurements at 1 Mc./s. , and a bridge method in the case of the other frequencies. The results for samples of barium and barium strontium titanate were in all cases analogous to those obtained at 1 kc./s. and 5 V./3 mm. ; within the limits examined the test frequency and tension had, therefore, no influence on the variations in permittivity observed.

The possibility was considered that the residual effects observed might be connected with the slow decay of space charges built up close to the electrodes of the samples. To test this explanation a high tension was applied to the dielectric for a short period, followed successively by further applications of lower tensions, sometimes in the same, sometimes in the opposite, direction to that of the initial field. The tensions were applied in such a way that the permittivity change due to the first treatment had not subsided entirely when the final tension was applied. The variations in permittivity due to these treatments for temperatures below and above the Curie point are shown in figs. 5 and 6, respectively. The permittivity change due to each treatment is seen to be increased by the following application of the tension, irrespective of the direction of the latter. The effects are, therefore, not due to space charges, since the application of a reverse field would then be expected to act in opposition to the previous treatment.

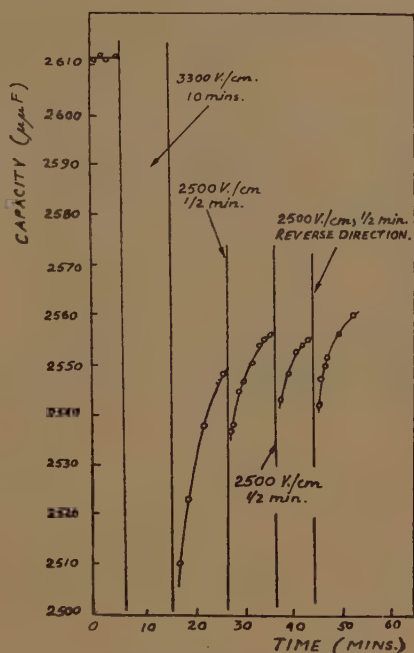
For one composition, prepared with the addition of cerium dioxide, a third effect was observed after the exposure to electric fields, in addition to the temporary variations in capacity obtained in all samples. The application of tensions exceeding 500 V./cm. resulted here in an apparently permanent decrease in the permittivity (fig. 7). On the application of several successive treatments, these permanent changes were found to become smaller, until finally a temporary increase only was observed.

Fig. 5.



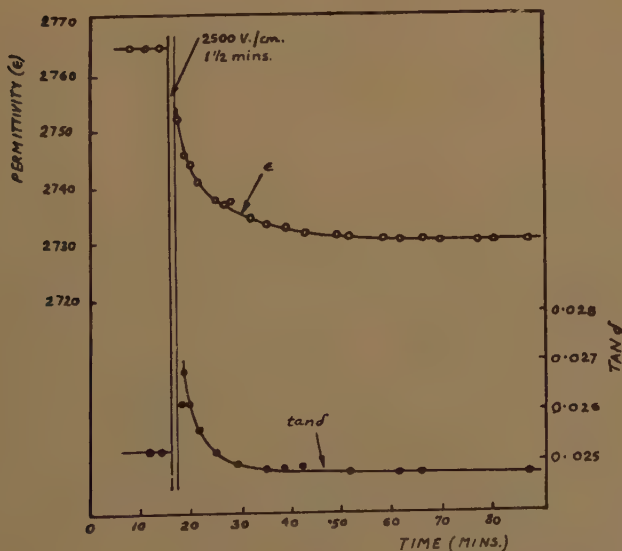
Sample A.2 [BaTiO_3].
Temperature 28°C .

Fig. 6.



Sample B.2 [$(\text{BaSr})\text{TiO}_3$].
Temperature 60°C .

Fig. 7.

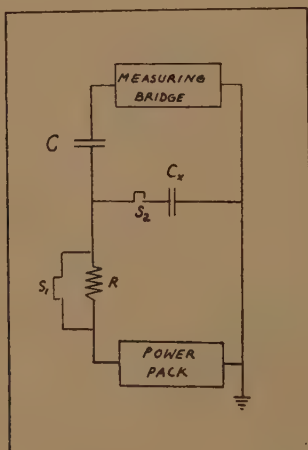


Sample C.1. D.C. Field 2500 V./cm., applied for $1\frac{1}{2}$ mins.
Temperature 26°C .

To ascertain the effect of the direction of the electric field, a sample of this type was subjected successively to two tensions applied in opposite directions, and the variations in capacity due to each treatment were followed in the usual way. The curves obtained were analogous to those in the case of successive tensions applied in the same direction.

The persistence of the modification in the value of the capacity over longer periods of time was next examined. Electrodes were applied to a specimen in such a way that one of the flat surfaces of the disk was coated completely with silver, while on the other the metal was applied in two separate portions. Electrical contact was made to the electrodes by wires soldered to each of the three coatings. A decrease of 3 per cent was effected in the permittivity of one of the sections of the disk by subjecting it to a high tension. The capacity was then measured at various time intervals and compared with that of the untreated portion.

Fig: 8



The difference in the values of the treated and untreated sections of the disk was found to remain effectively constant for the period of 14 days during which the measurements were carried out. The disks were stored over calcium chloride in a desiccator between the readings. On heating the sample to 100°C . the capacity of the treated portion reverted to its original value. A further application of a high tension resulted in a renewed decrease of the capacity which again reverted to its initial value on heating.

The permittivity of barium and barium strontium titanate, measured with a weak alternating field, is known to be decreased when a strong unidirectional electric field is superimposed upon the dielectric. Since no measurements appear to have been carried out of the variations in permittivity during the exposure to the field, it was considered of interest to determine the curves relating the permittivity to the time of application

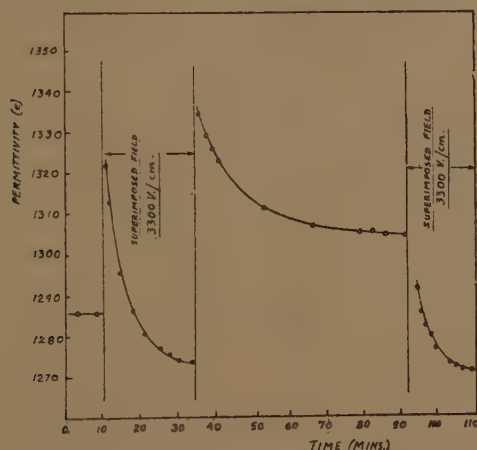
of a biasing tension. For this purpose the circuit arrangement shown in fig. 8 was set up. C_x represents the dielectric specimen which was contained in the air thermostat; C is a 1μ F mica blocking condenser and R a $10\text{ M}\Omega$ high-stability carbon resistor.

Before taking measurement, switch S_1 was closed to charge condenser C . The capacity of the specimens was generally between 1000 and $6000\mu\mu$ F. The effect of condenser C could, therefore, be neglected, since the total capacity measured, C_n , approximates to C_x when C is large compared with C_x :

$$C_n = \frac{CC_x}{C+C_x} \simeq C_x.$$

Thus a change as high as 10 per cent in the capacity of condenser C would result in a variation of only 0.04 per cent in the reading, C_n , if the capacity of the test specimen is $5000\mu\mu$ F.

Fig. 9.

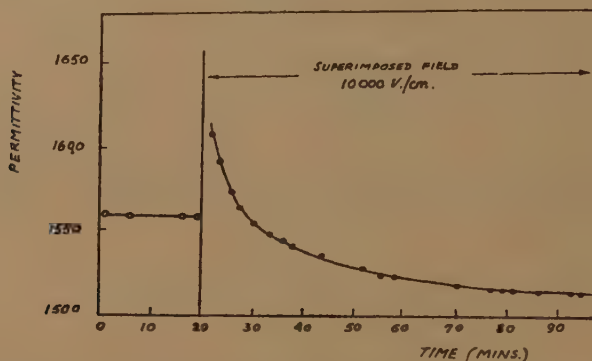


Sample A.3 [BaTiO_3]. Superimposed Field 33 00 V./cm. Temperature 32.5°C .

The arrangement was first of all tested by connecting a standard air condenser in place of the specimen, and measuring the capacities in the presence and absence of a biasing field. The value remained unchanged, showing the circuit to be satisfactory. Figs. 9 and 10 show typical measurements for barium titanate at temperatures below the Curie point. When a superimposed field is applied, the curves show that the permittivity has at first an increased value and then falls at a diminishing rate to a lower magnitude. Similar curves were obtained at various temperatures, with measuring tensions of 1 V./3 mm. and 5 V./3 mm., and frequencies of 500 c./s., 1 kc./s. and 15 kc./s. After removal of the D.C. field, condenser C was in all cases discharged before further measurements in the absence of the field were taken.

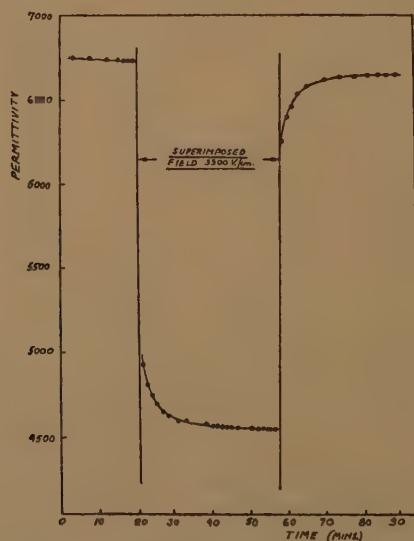
Fig. 11 shows the type of curve obtained at temperatures above the Curie point (sample B.2; barium strontium titanate). The permittivity here decreased immediately, approaching a steady reduced value. The effects observed below and above the Curie points were again seen to become greater on approaching the transition temperature. On releasing the biasing field the temporary variations described above were again

Fig. 10.



Sample A.2 [BaTiO_3]. Superimposed Field 10,000 V./cm. Temperature 27.5°C .

Fig. 11:



observed. These results, obtained with two different samples, were confirmed at temperatures below and above the Curie point in a specimen with a transition temperature of 50°C .

As a comparison, similar measurements were carried out on a silvered mica condenser having a capacity of the same order as that of the titanate specimens. No variations in the capacity could be detected in this case, either during or after the application of a unidirectional field.

DISCUSSION.

If the curve relating the permittivity to the temperature for specimens of barium titanate is considered, this is seen to show the characteristic sharp maximum at the transition point of the two crystalline structures (fig. 1). The loss tangent falls to relatively low values above the Curie point, but rises again rapidly at temperatures above 140°C. , owing to the increasing electrical conductivity of the material.

The curves showing the variations in permittivity after the application of an electric field reveal two new residual effects, observed in the case of all compositions examined. At temperatures below the Curie point, *i. e.* in the range corresponding to the distorted perovskite structure, the permittivity and loss tangent had at first an increased value, and reverted at a diminishing rate to their original magnitudes. Above the Curie point, in the region of the cubic structure, the permittivity was lower shortly after applying the tension, and reverted again slowly to its value prior to the treatment (Partington, Planer and Boswell 1947). The only apparent exceptions were specimens of barium strontium titanate at a temperature just above that of the permittivity maximum, $T_{\epsilon_{\text{max}}}$, a temporary increase in permittivity then being observed. It appears that the latter may be due to the co-existence of the two crystal structures within a region extending slightly beyond $T_{\epsilon_{\text{max}}}$, the effect due to the tetragonal modification being the larger one. At temperatures above the range of co-existence there was a temporary decrease in permittivity also, in this case, as in the other compositions.

Identical curves were obtained by the use of different alternating field strengths and frequencies, and these tests, therefore, show the results to be unaffected by the varying agitation produced in the dielectric by the measuring fields, within the limits examined.

The application of strong D.C. fields to certain types of dielectric materials results in the formation of space charges, the decay of which may take place over prolonged periods of time. In view of the difference in sign of the permittivity changes below and above the Curie point, and the increase in their magnitudes on approaching the latter, it was apparent that the above effects could not be caused directly by space charges. It was, on the other hand, thought possible that charges built up may in turn bring about a distortion in the crystal lattice, resulting in variations of the permittivity. The reversion of the permittivity value could in such a case be ascribed to the slow decay of the space charge.

Tests carried out with D.C. fields applied in opposite directions, however, showed that a mechanism of this type could not be responsible for the variations in permittivity observed (figs. 5 and 6).

In the case of the curves showing the changes in the permittivity during the superposition of an electric field, it is seen that these are not instantaneous on applying the tension, but that relatively slow variations take place. At temperatures below the Curie point, the permittivity exceeded at first that prior to the treatment, and fell gradually to a steady,

reduced value (figs. 9 and 10). Above the Curie point the permittivity decreased immediately, approaching similarly a final, steady value below that of its original magnitude (fig. 11) (Partington, Planer and Boswell 1948).

An explanation which appears to account adequately for the various effects observed may be suggested from the following considerations.

At temperatures below the Curie point, polycrystalline titanate compounds are composed of ferroelectric domains of random orientation. On the application of an external field the polarization of the material involves the following two mechanisms. In the first domains, the direction of which coincides with that of the field grow in volume at the expense of others, a displacement of domain boundaries taking place, reminiscent of the ferromagnetic case. In addition, orientation of domains may take place, involving probably a reversal in the polarity, equivalent to a rotation through an angle of 180° .

It was pointed out by Megaw that the action of anisotropic mechanical stresses on the tetragonal crystallites in barium titanate may be expected to result in a modification in the transition temperature between the tetragonal and cubic structures. The co-existence of the two forms in the temperature region of the permittivity maximum, has, for example, been attributed to the varying stresses on the crystallites in the material. It is interesting to note that a similar effect is observed also in the case of a λ -transition at -30°C . between two forms of ammonium chloride. The two phases co-exist in the region of the transformation, and the mechanism is also believed to involve displacements of the transition point, owing to local stresses on the crystallites.

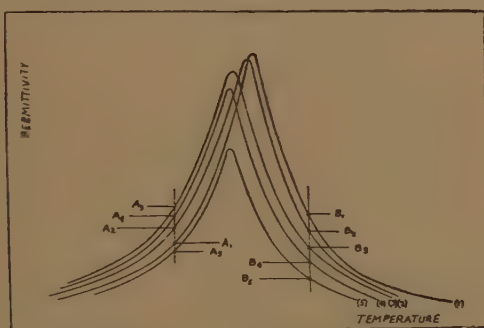
It appears that the following mechanism may be responsible for the residual effects observed. On the application of the high tension, regions in the material the direction of which coincides with that of the polarizing field increase in volume at the expense of others. Owing to these changes, anisotropic stresses are produced on the crystallites, resulting in the displacement of the Curie point to a lower temperature. In the case of the tetragonal structure, for example, compressive forces in a direction parallel to the c -axis may predominate over those along the a -axis, so facilitating a transition to the cubic modification. At temperatures below the Curie point this "grain-growth" will involve the displacement of domain boundaries, molecules being thus transferred to the aligned regions. Orientation of domains may, in addition, take place. Since the ferroelectric influence extends to temperatures above the Curie point, it is reasonable to assume that a similar growth of oriented regions may also take place in a strong field somewhat above the transition temperature. Simultaneously with the displacement of the Curie point, the permittivity-temperature curve will be lowered progressively in a direction parallel to the permittivity axis, owing to the increasing polarization of the material.

On removal of the field, molecules added to the oriented regions revert to their initial state, and the Curie point assumes its original value. At

the same time, the permittivity-temperature curve is raised again in the direction of the permittivity axis. Changes in volume of the oriented regions, while rapid immediately on applying or removing the electric field, may continue to take place for relatively long periods of time.

The proposed displacements of the permittivity-temperature curve are indicated in fig 12. Curve (1) represents the position prior to the treatment. On applying a unidirectional field, the curve is displaced progressively from (1) to (5); on releasing the tension, it reverts from (5) through the stages (4), (3) and (2) to its initial position (1). When a biasing field is applied, the permittivity A_1 at a temperature below the Curie point therefore passes successively through the stages A_2 , A_3 and A_4 , assuming finally a value A_5 . On removal of the tension, the permittivity reverts from A_5 to its original magnitude A_1 . Similarly, the permittivity B_1 at a temperature above the Curie point falls to B_5 on applying the

Fig. 12.



field, and reverts to B_1 after its removal. Fig. 13 shows the resulting variations according to this scheme, at temperatures below and above the Curie point. In the former case the curve attains an initial maximum and falls subsequently to a low value of the permittivity, during exposure to the field. On releasing the tension the curve rises again to the same maximum, and then approaches the initial permittivity value. Above the Curie point the curve falls immediately on applying the field and reverts to the original permittivity on its removal.

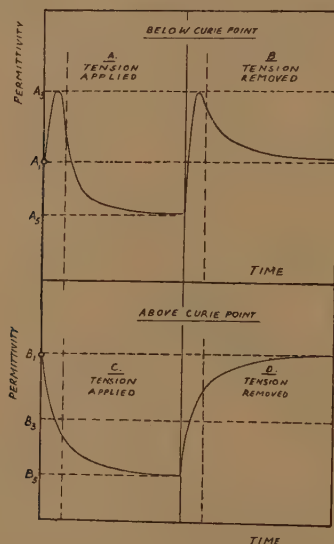
Since the boundary displacements of the oriented regions are greatest immediately on the application or the removal of the field, the initial variations of the curves in fig. 13 are too rapid to be recorded. The vertical broken lines indicate the times at which the permittivity measurements were begun. Fig. 13 (A to D) corresponds, therefore, to the experimental curves actually obtained during, as well as after the application of D.C. fields, both below and above the Curie point. Since the displacement of the permittivity-temperature curve is dependent on the total volume of the regions aligned parallel to the applied field, the above explanation is consistent also with the result obtained on the application

of successive tensions of opposed directions. A mechanism such as that suggested will, therefore, explain the various residual effects observed.

The variations in the loss tangents observed after removal of the field appear to indicate that the dielectric loss of the partially polarized material, the crystallites of which are acted upon by local stresses, exceeds that in the unpolarized dielectric. The reversion to random orientation is, therefore, accompanied by a gradual fall in the loss tangent to its value prior to the treatment.

It is seen from the magnitude of the permittivity changes observed that the displacement of the Curie point resulting from the polarizing

Fig. 13.



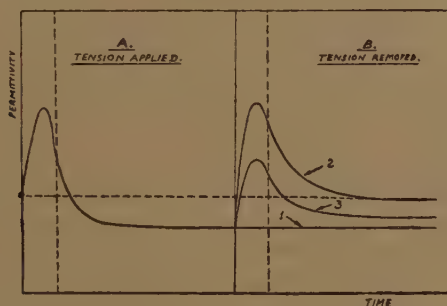
field is relatively small. This cannot, therefore, be detected by determining the permittivity-temperature curve in the presence of a biasing field, particularly in view of the uncertainty introduced into the measurement of the permittivity maximum by the co-existence of the two modifications of the material in this region.

The processes taking place during and after exposure to the field are seen from the experimental curves to be unsymmetrical, the permittivity reverting only slowly to its original value after relatively short applications of the tension. This behaviour is, however, shown also by other ferro-electric materials, prolonged periods of depolarization being, for example, observed in the case of Rochelle salt. The reversion of the permittivity values observed in titanates on removal of the tension, being the resultant of a displacement of the permittivity-temperature curve in two directions, is found not to follow an exponential law. This may be seen from a plot of the logarithms of the permittivity values against the time, which shows that the relation is non-linear.

It is interesting to recall that the permittivity maxima in barium titanate have been reported to occur at lower temperatures in the case of strong A.C. measuring fields than at lower field strengths (de Bretteville 1946, Jonker and van Santen 1947). It is possible that this may similarly be due to internal mechanical stresses, caused here by the alternating measuring field. The mechanism suggested above for the residual effects is consistent with this result.

A similar explanation for the effects observed by us has been suggested by Szigeti. According to this, the variations in permittivity may result from a displacement of the permittivity-temperature curve of the dielectric due to the field, along the temperature axis. The permittivity maximum is assumed to be shifted to a lower temperature, but to retain its original magnitude. This explanation, however, would not in itself be sufficient to account for the prolonged residual effects observed, and it seems necessary, in addition, to take into account the depression in the permittivity-temperature curve, as well as some mechanism for the displacement of the Curie point such as that suggested above.

Fig. 14.



The permanent decrease in permittivity observed at high field strengths in samples prepared with cerium dioxide may be explained on lines similar to those proposed above. Here, the reversion to the initial state of a part of the domains enlarged by the polarizing field is prevented or greatly retarded. On removal of the tension, molecules added to these aligned domains, therefore, remain incorporated in the latter. Since this results in an increase in volume of the domains oriented in the direction of the alternating measuring field, the overall permittivity will fall to a lower value. In fig. 14 B, curves (1) and (2) show the proposed permittivity-time variations corresponding to the two processes taking place simultaneously on cessation of the tensions; (1) shows the curve for the domains which retain their enlarged volume, and (2) that for domains reverting to their initial state. The resultant permittivity varies, therefore, as indicated by curve (3). This corresponds with the experimental result. Fig. 14 A shows the variations in permittivity during exposure to the field.

On following the high-tension treatment with a second application of a field opposed in direction to the initial one, a further decrease in permittivity was observed. According to the above explanation, the second treatment results in the growth of domains opposed in polarity to those the volume of which was increased by the initial field. Although the polarization of the specimen is, therefore, reduced by the second treatment, the total volume of domains aligned in the direction of the measuring field increases. A fall in the permittivity is hence observed when measurements are carried out with an alternating field. Warming of the dielectric may here facilitate the exchange of molecules between neighbouring domains, and at higher temperatures the sample, on removal of the tension, reverts to its random state, as in the case of the previous specimens. It seems, therefore, that the permittivity variations in this type of sample may be accounted for by a mechanism similar to that suggested for the residual effects observed in the other compositions.

SUMMARY.

A study of dielectric residual effects in barium titanate and similar compounds possessing structures of the perovskite type has been made. Curves relating the permittivity to the temperatures were determined for several different compositions. The variations in permittivity during, as well as after the application of strong unidirectional electric fields, were studied at temperatures below and above the Curie points. Several new effects were revealed by the permittivity-time curves so obtained. A temporary increase in permittivity resulted after the application of the field at temperatures below the Curie point, while above the latter, a temporary decrease took place. In the case of one type of composition a permanent decrease in permittivity was obtained under certain conditions. During the application of a superimposed field below the Curie point, the permittivity had at first an increased value and fell gradually to one below its initial magnitude. Above the Curie point it decreased immediately, approaching again a reduced value.

A mechanism is proposed for the various effects observed, involving a displacement of the Curie point to a lower temperature on the application of the field, accompanied by a depression of the permittivity-temperature curve along the permittivity axis. The displacement of the Curie point is ascribed to the growth of oriented regions in the material, resulting in anisotropic stresses on the crystallites.

ACKNOWLEDGMENTS.

We are greatly indebted to the United Insulator Co., Ltd., Surbiton, for generously giving us facilities in their Research Laboratory during part of this investigation. We are grateful also to Dr. B. Szigeti for helpful suggestions.

REFERENCES.

- CADY, 1946, *Piezoelectricity*, chap. 22. (McGraw-Hill Book Co. Inc.).
DE BRETTEVILLE, 1946, *J. Amer. Cer. Soc.*, **29**, 303.
JONKER, and VAN SANTEN, 1947, *Chemisch. Weekbl.*, **43**, 672.
MEGAW, 1946, *Trans. Faraday Soc.*, **42**, 224; 1947, *Proc. Roy. Soc. A*, **189**, 261.
PARTINGTON, PLANER, and BOSWELL, 1947, *Nature, Lond.*, **160**, 877; 1948, *Ibid.*, **162**, 151.
REDDISH, PLESSNER, and WILLIS-JACKSON, 1946, *Trans. Faraday Soc.*, **42**, 244.
ROBERTS, 1947, *Phys. Rev.*, **71**, 890.
RUSHMAN, and STRIVENS, 1946, *Trans. Faraday Soc.*, **42**, 231; 1947, *Proc. Phys. Soc.*, **59**, 1011.
SZIGETI, private communication.
VALASEK, 1922, *Phys. Rev.*, **19**, 478.
WUL, 1946, *J. Physics, U.S.S.R.*, **10**, 95.
-

XIV. *On the Oxidation of Metals at Low Temperatures and the Influence of Light.*

By N. CABRERA,

H. H. Wills Physical Laboratory, University of Bristol*.

[Received September 30, 1948.]

§1. INTRODUCTION.

THE oxide layers formed on many metals, when they are exposed to an oxygen atmosphere, have a compact structure. Assuming that this structure does not differ very much from that of the perfect oxide crystal, it has been possible to develop a theory of oxide layers which seems to account for a number of experimental facts. The purpose of this paper is to develop this theory in several respects. As indicated, we shall neglect from the beginning the effects produced by the presence of discontinuities (intercrystalline surfaces, dislocations, etc.) in the oxide lattice.

At very high temperatures (say 1000° C.) the law of growth is usually parabolic. The theory of the oxidation at these temperatures, developed particularly by Wagner (1933), is based on the hypothesis of the diffusion of electrons and metallic ions through the oxide, their distribution being such that the space charge is always neutralized.

At room temperatures the growth of the oxide layer follows a logarithmic type of law. This behaviour has been studied particularly for aluminium, but it occurs also in a number of other metals: zinc, copper, chromium, tungsten, etc., and seems to be very general. It is difficult to believe that at these temperatures the metallic ions can diffuse through the oxide without the help of an external force (electric field for example). This can occur by a mechanism suggested by Mott (1947). Mott first suggested (1940) that at low temperatures the metallic ions and electrons could be considered independently, because the space charge set up is then very

* Communicated by Prof. N. F. Mott, F.R.S.

small. This can be seen as follows. Let us suppose that the ions diffuse via the interstitial positions and the electrons via the conduction band of the oxide. Then, using the methods developed by Debye and Hückel in their theory of electrolytes, it is possible to show that the concentration of ions and electrons neutralize each other only for distances bigger than the Debye length x_0 ; given by

$$x_0^2 = \kappa kT / 2\pi N e^2, \quad (1)$$

where κ is the dielectric constant of the oxide and N the concentration of electrons and ions, where the space charge is zero. Putting numerical values in (1) one obtains

$$x_0 \sim 10^{-8} \exp \{(W + s\phi) / 2(s+1)kT\} \text{ cm.}, \quad (2)$$

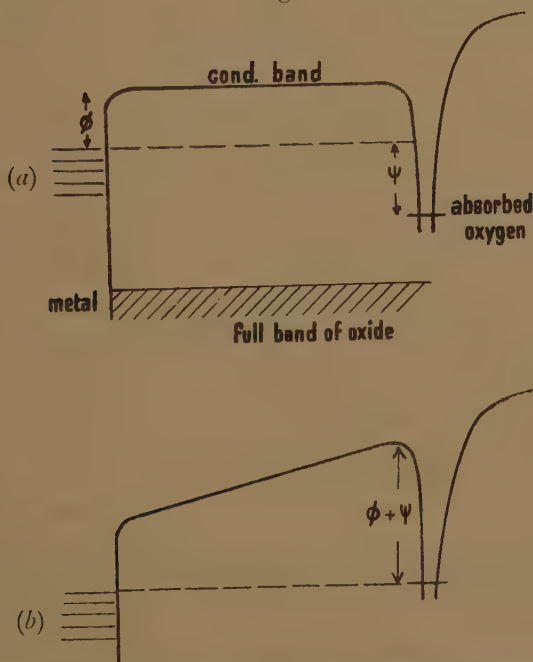
where W is the heat of solution of metal ions (valency s) in the oxide and ϕ the work function of the electrons from the metal to the oxide. For reasonable values of W and ϕ , x_0 is c. 1 cm. at room temperatures and decreases till c. 10^{-6} cm. at 1000°C . On the contrary, the thickness of oxide observed increase from c. 10^{-7} cm. at room temperatures to c. 10^{-1} cm. at 1000°C . Therefore, it is clear that for temperatures below 600°K . we should not expect to find any influence of the space charge on the diffusion of ions and electrons; they can therefore be treated separately. On the contrary, above 1000°K ., the influence of the space charge is so big that the ions and electrons must diffuse in such a way that the space charge is neutralized. The intermediate region has not yet been studied theoretically.

For low temperatures, for which we can neglect the effect of the space charge, Mott (1947), has put forward the following picture: he assumes that there is adsorbed oxygen on the oxide-air surface of the oxide layer, and also that the electrons can pass easily from the metal into the empty levels of the adsorbed oxygen atoms or molecules. Then an electric field is set up through the oxide layer strong enough to produce a stationary current of ions even if their diffusion were impossible in the absence of the field. We shall discuss this picture in the following section (§ 2), and consider its extension to oxides, such as Cu_2O , for which the metallic ions diffuse by the mechanism of metallic vacant lattice sites created on the oxide-air surface; we shall also study under what conditions (low pressures, very low temperatures) the controlling factor in the growth may be the reaction on the oxide-air surface, which will give a linear law of growth.

Mott (1940), proposed another theory (now abandoned) in order to explain the oxidation at low temperatures. He supposed that the ions could diffuse easily through the oxide at these temperatures, but the work function ϕ (metal-oxide) for the electrons is too high for them to go into the oxide. Under these conditions we should expect to have a logarithmic law also, because for very small thicknesses ($< 40 \text{ \AA}$) of oxide we know that the electrons can cross the layer by quantum-mechanical tunnel effect. If this picture were correct, the author (1945)

suggested that light of a frequency ν , such that $h\nu > \phi$, should increase considerably the rate of growth of the oxide layer. The experiment was made on aluminium by Cabrera, Terrien and Hamon (1947), and the result was that the thickness of oxide increased only by a factor of about 2, which is much smaller than we should expect if the 1940 picture were correct. On the other hand, we know now that this is not so for several reasons, in fact the ions cannot diffuse at room temperature through most oxides, otherwise they would be ionic conductors at these temperatures, which is not the case. Now it is necessary to explain, on the basis of the 1947 picture, the fact that there is a small influence of light on the oxidation of aluminium. In section 3 of this paper we shall give a theory of this effect, which seems to be in accord with the experimental results.

Fig. 1.



Electronic levels in the metal, oxide and adsorbed oxygen: (a) before the equilibrium is reached; (b) after.

§ 2. OXIDATION OF METALS AT LOW TEMPERATURES.

1. Let us consider first what happens when there is a large concentration of adsorbed oxygen molecules and atoms on the oxide-air surface in equilibrium with the atmosphere.

Following Mott (1940, 1947) we suppose that the empty levels of the adsorbed oxygen are below the top of the Fermi distribution in the metal; let us call ψ the difference of energy between the Fermi distribution and the adsorbed oxygen levels (fig. 1 (a)).

(i) At very low temperatures the solubility and the mobility of the metallic ions into the oxide is negligible. If there is an oxide layer formed on the metal, it is due exclusively to the fact that the thickness x is then very small and therefore the field is very big (3). Consequently we have to consider only the metallic ions going in the direction metal-oxide.

Let us now suppose that the activation energy necessary to take one ion from the *metal surface* and put it into the first interstitial position of the oxide U (fig. 2) is the same for all the surface ions. It must be mentioned that this hypothesis is not correct. The energy in question is different for different positions on the metal surface and for different crystallographic surfaces. We shall study this point in another paper, and we shall prove that the hypothesis above is a good approximation for very fine polycrystalline materials to which we refer in this paper.

Under these conditions the current of ions is given by the expression

$$j_m = \frac{\nu}{a^2} e^{-U/kT} e^{e_i a F / kT}, \quad (4)$$

where $a^{-2} \sim 10^{57} \text{ cm.}^{-2}$ is the number of metallic ions per cm.^2 of the metal surface and $\nu \sim 10^{12} \text{ sec.}^{-1}$ the frequency of atomic vibrations. Then, using (3), the rate of growth of the oxide layer is

$$\frac{dx}{dt} = j_m \Omega = a\nu \exp \left\{ - \left[U - \frac{e_i a \psi}{ex} \right] / kT \right\}, \quad . . . (5)$$

where $\Omega \sim a^3$ is the volume of oxide per metallic ion. This formula gives a very rapid growth for small x , but growth ceases when x is so big that $(U - e_i a \psi / ex) \gg kT$. If x_c is the limiting thickness for which the rate of growth is very small (say $x_c = a \times 10^{-7} \text{ cm. sec.}^{-1}$), Mott deduces from (5) the formula

$$\frac{x_c}{a} = \frac{C}{1 - (T/T_c)}; \quad C = \frac{e_i \psi}{e U}; \quad T_c = \frac{U}{45k}. \quad (6)$$

(ii). For $T > T_c$ there is no limiting thickness; in fact, at these temperatures x is not small, and the field F is not large, so in calculating the current we have to subtract from (4) the contribution given by the ions going in the direction oxide-metal. Mott obtains in this way the general formula for the current

$$j_m = \frac{\nu}{a^2} e^{-U/kT} 2 \sinh \left(\frac{e_i a F}{kT} \right), \quad (7)$$

which reduces to (4) when $e_i a F \gg kT$. For high temperatures (7) gives a rapid initial rate of growth followed by a parabolic one, without limiting thickness. Formula (7) will remain valid at temperatures and thicknesses for which the space charge in the oxide remain small (§ 1).

The temperature T_c , for which the law of growth becomes parabolic, is proportional to U . In the case of aluminium oxide, Mott (1948) modifying earlier calculations of Verwey (1935), determined U and a by

comparing the equation (4) with the experimental exponential law, obtained by Günterschulze and Betz (1934), for the ionic current as a function of the field, during the anodic oxidation of aluminium in borax solutions at 20° C. The values found are $U=1.8$ eV. and $a=3.5$ Å. From the value for U , Mott obtains $T_c \sim 200^\circ$ C., which seems to be confirmed by the experiments of Cabrera and Hamon (1947) (the experimental value seems to be between 200° and 300° C.). The limiting thickness x_c is, on the other hand, proportional to $a\psi$; in the case of aluminium again, taking $a=3.5$ eV. and assuming $\psi \sim 2$ eV., one obtains the right order of magnitude.

It should be mentioned that in the case of the oxides for which the metal ions diffuse via the interstitial positions of the oxide, the growth of the oxide layer is independent of the pressure of oxygen, provided it is high enough for there to be sufficient absorbed oxygen ions to maintain the electronic equilibrium described above (see § 22). For the same reason the growth of the oxide layer is also independent of the presence of ozone in the atmosphere, as has been shown experimentally by Cabrera, Terrien and Hamon (1947).

An analogous theory can be developed for oxides in which the metallic ions diffuse by the mechanism of vacant lattice sites, formed on the oxide-air surface (case of copper), under the influence of the same field (3). Campbell and Thomas (1947) have shown that the growth of oxide layers on copper between 100° and 250° C. follows a cubic law and not the parabolic law (7) observed in the case of aluminium between 300 and 500° C. (Gulbransen and Wyson 1947). Probably the explanation of this result resides in the fact that the number of vacant sites created on the oxide-air surface must be related to the concentration of adsorbed oxygen ions in equilibrium with the metal, according to Mott's picture, because the energy required for the formation of a vacant site near an adsorbed oxygen ion must be smaller than that required near an absorbed oxygen atom.

The number of adsorbed oxygen ions n per cm.² is related to the field F by the Coulomb formula

$$F = \frac{4\pi ne}{\kappa}, \quad (8)$$

which by comparison with (3) gives

$$n = \frac{\kappa\psi}{4\pi e^2 x}, \quad (9)$$

showing that n is inversely proportional to the thickness. If we assume that the holes are *only* formed in the neighbourhood of an adsorbed oxygen ion, requiring an activation energy U , the current should be given by the formula

$$j_m = nve^{-U/kT} \sinh \left(\frac{e_i a F}{kT} \right), \quad (10)$$

which for high temperatures ($e_i a F/kT \ll 1$) should give a cubic law of

growth. At low temperatures (the critical temperature T_c seems to be between 50°C . and 100°C . for copper), we should expect a logarithmic law as Lustman and Mehl (1941) have observed.

It should be mentioned that formula (10) gives a law of growth independent of pressure at all temperatures, provided the pressure is high enough to maintain Mott's electronic equilibrium and the films are thin enough to avoid space charge effects. For thick films and high temperatures, of course, the rate of growth becomes proportional to p^{18} in the case of copper, as was proved by Wagner.

2. We consider now what happens when the rate of supply of oxygen from the air is not large enough to maintain the equilibrium required by Mott's picture.

When there is enough oxygen, then first there is established an equilibrium concentration of adsorbed oxygen depending on the pressure, and secondly, an equilibrium concentration of adsorbed oxygen ions, smaller than that of adsorbed atoms or molecules, and given by formula (9). Now n increases very rapidly when x decreases; consequently, there must always be a thickness below which the supply of oxygen is not big enough to maintain this equilibrium concentration. Then the rate of growth of the oxide will be proportional to the number j_0 of oxygen ions produced per sec. per cm^2 at the oxide-air surface, and consequently the law of growth will be linear.

For a given j_0 , the linear law will go over into Mott's law at the thickness for which both rates of growth are equal. If j_0 is big, the linear law will be always confined to the initial very rapid growth of Mott's law and therefore will be unobservable.

The value of j_0 will depend on the mechanism transforming the oxygen molecules of the air into O^{--} ions of the oxide lattice, a mechanism about which we do not know very much. Probably an adsorbed oxygen molecule is just ionized forming an adsorbed O_2^- ion, then it attracts a metallic ion and becomes dissociated without the need of an activation energy. Therefore, under the conditions of the linear law, we can suppose that the oxygen molecules striking the oxide surface are almost immediately transformed into oxygen ions of the lattice; consequently the current j_0 is simply given by

$$j_0 \sim p / \sqrt{2\pi M k T}, \quad \dots \dots \dots (11)$$

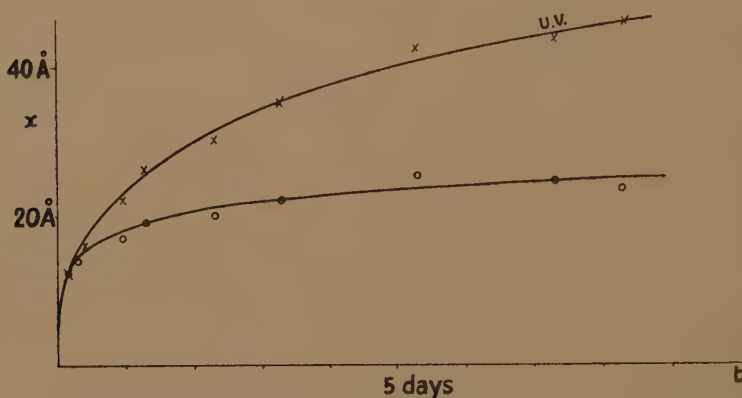
where p is the pressure and M the mass of an oxygen molecule. For high values of p the linear law will be unobservable; only when the rate of growth $j_0 \Omega$ is smaller than, say, $50 \text{ \AA}/\text{hour} \sim 10^{-10} \text{ cm. sec.}^{-1}$, we shall be able to observe it. From (11), assuming $\Omega \sim 10^{-23} \text{ cm}^3$, this corresponds to pressures such that

$$p < 10^{-4} \text{ mm. Hg.} \quad \dots \dots \dots (12)$$

We deduce, therefore, that *Mott's law, always independent of pressure, will be valid down to 0°K . and for pressures above $\sim 10^{-4} \text{ mm. Hg}$. This limiting pressure will be practically independent of temperature and of the oxide considered.*

It could perhaps happen that some activation energy V is required for the dissociation of the O_2^- adsorbed ions; then the current j_0 should be an exponential function of the temperature, and below a temperature of the order $V/35k$ we should observe only the linear law, given a very small thickness of oxide. This is certainly not so for oxides for which the metal diffuses via the interstitial positions, because the heat of formation of the oxide is then released on the oxide-air surface and can be used to overcome any activation energy possibly existing at this surface. If the metal diffuses by the mechanism of vacant lattice sites, the heat of formation of the oxide is released on the metal-oxide surface, and therefore cannot be used to overcome the activation energy V , if it exists. Measurements at very low temperatures could decide this point.

Fig. 3.



Influence of ultraviolet light on the oxidation of aluminium.

§ 3. INFLUENCE OF LIGHT ON THE OXIDATION OF ALUMINIUM.

Cabrera, Terrien and Hamon (1947) have studied recently the influence of ultraviolet light on the rate of oxidation of aluminium at room temperature. The experimental results are represented in fig. 3; the upper curve corresponds to the oxidation under the influence of ultraviolet light, and the lower curve gives the law of growth under ordinary conditions. The source of light was a mercury-arc, giving an intensity of the order of 10^5 erg/cm.² sec. for a frequency ν , such that $h\nu \sim 4$ eV.

The following hypothesis has been made to account for the effect* :

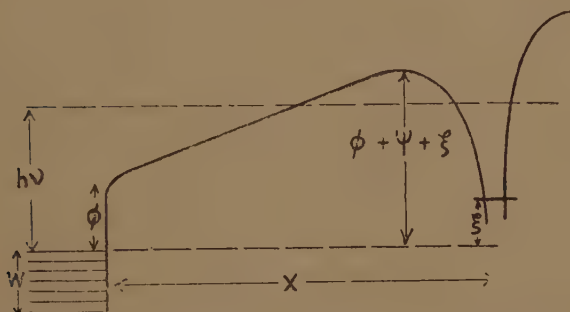
Under the influence of the light the electronic equilibrium represented by fig. 1 (b) is disturbed, the number of electrons coming from the metal being increased. Consequently, in order to obtain a new equilibrium we must increase the field existing in the oxide so as to hinder diffusion of the electrons coming from the metal more difficult, and make the diffusion of the electrons going in the other sense easier. This means, of course, that the levels in the adsorbed oxygen atoms will now be raised above the top of the Fermi distribution in the metal; let us call this

* The author is indebted to Prof. N. F. Mott for suggesting this picture.

rise ξ (fig. 4). Now an increase of the field inside the oxide means of course an increase of the rate of growth, which is actually observed.

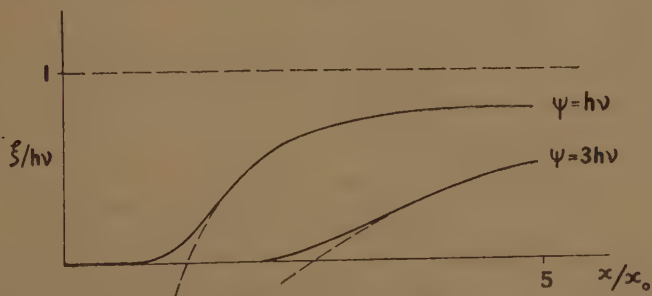
Actually it is easy to see that the value of ξ will depend on the thickness x of the oxide layer. For small values of x (several atomic layers of oxide) the electrons can pass very easily through the oxide layer by tunnel effect, consequently the light will increase the current of electrons going in *both* directions, but the same equilibrium represented by fig. 1 (b) will exist. Therefore we expect to have $\xi \rightarrow 0$ for $x \rightarrow 0$. As the thickness increases the transmission coefficient of the electrons through the oxide

Fig. 4.



Influence of light on the electronic equilibrium set up in the oxide layer.

Fig. 5.



$\xi/h\nu$ in function of x/x_0 for different $\psi/h\nu$ (formula (22)). The constant x_0 is $\sim 20\text{\AA}$ for Al.

becomes rapidly smaller. Then the increase of the electronic current coming from the metal cannot be compensated by an increase of the opposing current if $\xi=0$ (fig. 4). In order to obtain a new equilibrium we must increase ξ . In this way we increase the electronic current in the direction oxide-metal and we decrease also the current in the other direction because only the electrons having (after absorption of a quantum) an energy between $W+\xi$ and $W+h\nu$ (fig. 4) can pass through the oxide. We see that ξ increases with x , but its maximum value must be $h\nu$. We see that ξ increases with x , but its maximum value must be $h\nu$. Therefore we expect a variation of ξ with x , such as is illustrated in fig. 5. We have deduced this general behaviour of $\xi(x)$,

assuming that the mean free path of the electrons in the oxide is bigger than the thickness of the oxide layer. This hypothesis seems reasonable for Al_2O_3 and the thicknesses obtained at room temperature; in any case we do not think that the form of $\xi(x)$ will change very much if that is not entirely correct.

We can see at once that these qualitative considerations are enough to explain the experimental results. Experimentally there is no influence of light until thicknesses bigger than $\sim 10 \text{ \AA}$, then the influence increases steadily until the thickness is multiplied by a factor of the order 2. Now we know that the thickness is proportional to ψ ; therefore, in the presence of light, we expect to have a thickness proportional to $\psi + \xi\nu < \psi + h\nu$; consequently we expect the thickness to be, under the conditions of the experiments quoted above, multiplied by a factor < 4 .

Let us now make some calculations in order to deduce the form of ξ , always assuming the mean free path of the electrons bigger than x . It is possible to make approximate calculations, when the thickness x is big enough for the transmission coefficient of the electrons through the oxide to be small. Then the important factor in the determination of ξ is this coefficient, and all the other factors can be estimated approximately.

We shall neglect the absorption of light in the oxide itself, which is reasonable for frequencies below the absorption band of the oxide.

We consider first the electronic current \vec{n} in the direction metal-oxide. The current of electrons in the oxide with energies ζ in the direction normal to the surface, in the interval $d\zeta$, is given, at least approximately, by the product of the following factors:—

(i) The number of metallic electrons, with normal energies ζ between ζ and $\zeta + d\zeta$, striking the metal-oxide surface per cm^2 per sec. This is given by the well-known formula

$$\frac{4\pi m}{h^3} (W - \zeta) d\zeta, \quad (13)$$

where we count the energy from the bottom of the Fermi distribution in the metal (fig. 4) and we neglect the influence of the temperature.

(ii) The probability for one of these electrons to make a transition to the state $\zeta' = \zeta + h\nu$, under the influence of the light of frequency ν . This probability cannot be calculated directly, because strictly speaking the separation into the factors which we are considering is not allowed. Nevertheless, Mitchell (1934) has calculated, by exact methods, the photoelectric emission from a clean metallic surface, assuming an ordinary step potential barrier (0 for $x < 0$, $W + \phi$ for $x > 0$), and has proved that the result can be factorized in the way considered here. The factor representing the excitation probability is, neglecting factors of the order 1,

$$\frac{I\alpha}{m\nu^3} f(\zeta, \zeta'), \quad (14)$$

where I is the intensity of light, $\alpha=e^2/c$, and $f(\zeta, \zeta')$ is a dimensionless quantity, depending essentially on the form of the potential barrier. We can expect that, at least approximately, the same formula should be applicable to our problem, but with a different value for $f(\zeta, \zeta')$, which is probably a small power of 10^* .

(iii) The transmission factor D for electrons, with energy $\zeta'=\zeta+h\nu$, through the potential barrier represented in fig. 4. D will increase from zero, for $\zeta'<W+\xi$, up to 1 for $\zeta'>W+\phi+\psi+\xi$. For large thicknesses, and for electrons with energies near $W+h\nu$ (fig. 4), the transmission coefficient is given by

$$D(\zeta')\sim\exp\left\{-\frac{4\kappa}{3eF}(A-\zeta')^{3/2}\right\}, \quad . \quad . \quad . \quad . \quad (15)$$

where $\kappa^2=8\pi^2m/\hbar^2$, $eF=(\psi+\xi)/x$ and $A=W+\phi+\psi+\xi$.

(iv) Finally we have to multiply by another factor which will take account of the fact that the electrons of the metal cannot go into the adsorbed layer if there is no empty level available. Let us assume that this factor is of the order of

$$(n_0-n)a^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

where n_0 is the total number of levels per cm^2 in the adsorbed oxygen, n the actual number of electrons per cm^2 in these levels, and a^2 the surface per level (of the order of the surface per atom). We suppose of course that there are always empty levels in the adsorbed layer, otherwise the law of growth in the absence of light would be linear (§ 2).

The total current of electrons in the direction metal-oxide is then

$$\vec{n}\sim a^2(n_0-n)4\pi\frac{I\alpha}{(\hbar\nu)^3}\int_0^W(W-\zeta)\cdot f(\zeta, \zeta')\cdot D(\zeta')\cdot d\zeta. \quad . \quad . \quad (17)$$

For thicknesses big enough, we see that the important factor here is D , and a rough evaluation only of f is necessary. The value of the integral in (17) is determined essentially by the values of ζ near their maximum W . Using (15) one obtains

$$\vec{n}\sim a^2(n_0-n)\pi\frac{I\alpha}{(\hbar\nu)^3}f_W\frac{(eF)^2}{\kappa^2M}\exp\left\{-\frac{4\kappa}{3eF}M^{3/2}\right\}, \quad . \quad . \quad . \quad (18)$$

where $M=\phi+\psi-(\hbar\nu-\xi)$ and $f_W=f(W, W+h\nu)$.

Let us now consider the current of electrons \vec{n} in the reverse direction. There are two mechanisms possible. The electrons can go back to the metal directly by tunnel effect through the oxide layer, or they can be excited first under the influence of the light; it can be shown that for reasonable values of the intensity the second mechanism gives a small contribution because the concentration of electrons in the adsorbed layer is small.

* The value of f in Mitchell's calculations is $\frac{W+\phi}{\hbar\nu}\sqrt{\frac{\xi}{\xi+h\nu}}\sim 1$.

The current produced by tunnel effect will be given, at least approximately, by the product of the following factors :

(i) The number of electrons per cm^2 in the levels of the adsorbed layer : n .

(ii) The number of times per sec. that one of these electrons is directed towards the metal ; we can estimate it to be of the order of the frequency corresponding to one of the adsorbed oxygen levels : $\nu_a \sim 10^{15} \text{ sec.}^{-1}$.

(iii) The transmission factor for these electrons (energy $\zeta = W + \xi$) through the oxide : $D(W + \xi)$.

The current \hat{n} is therefore given by

$$\hat{n} \sim n \nu_a \exp \left\{ -\frac{4\kappa}{3eF} N^{3/2} \right\}; \quad N = \phi + \psi. \quad . \quad . \quad . \quad (19)$$

The value of ξ will be determined now by the condition of equilibrium : $\hat{n} = \bar{n}$. Using (18) and (19) we obtain the equation

$$\frac{n}{n_0 - n} = \frac{I}{G} \exp \left\{ -\frac{4\kappa}{3eF} (M^{3/2} - N^{3/2}) \right\},$$

$$G = \frac{\nu_a (\hbar\nu)^3}{a^2 \pi \alpha f_w} \frac{\kappa^2 M}{(eF)^2}.$$

The factor G turns out to be very big, of the order of 10^{20} (see below).

Taking logarithms, and neglecting $\log \frac{n}{n_0 - n}$, we deduce

$$\frac{3}{2} \frac{\psi + \xi}{\phi + \psi} \frac{x_0}{x} = 1 - \left(1 - \frac{\hbar\nu - \xi}{\phi + \psi} \right)^{3/2}, \quad . \quad . \quad . \quad . \quad (20)$$

where the constant x_0 is given by

$$x_0 = \frac{\log(G/I)}{2\kappa(\phi + \psi)^{1/2}} \sim \frac{\log(G/I)}{(\phi + \psi)^{1/2}} \text{ \AA}, \quad . \quad . \quad . \quad . \quad (21)$$

if $(\phi + \psi)$ is written in eV. The maximum value for ξ given by the equation (20) is of course $\hbar\nu$. For small values of $(\hbar\nu - \xi)$ we obtain

$$\frac{\xi}{\hbar\nu} = 1 - \frac{\psi + \hbar\nu}{\hbar\nu} \frac{x_0}{x + x_0}. \quad . \quad . \quad . \quad . \quad (22)$$

This formula is plotted in fig. 5 for several values of $\psi/\hbar\nu$. It will be approximately valid for $x > x_0\psi/\hbar\nu$, where ξ given by (22) is positive. Below $x_0\psi/\hbar\nu$, the actual function ξ must become zero as indicated in fig. 5.

The influence of the intensity I on ξ is very small ; as the intensity increases the value of x_0 decreases very slowly. On the other hand, the influence of the frequency ν is fundamental. Apart from very small changes in x_0 , we see that ξ will be small for small ν , first because the

maximum of ξ is small and secondly because the distance at which it becomes sensibly different from zero ($x_0\psi/h\nu$) is big (fig. 5). The influence of the light on the oxidation will become observable for frequencies for which $x_0\psi/h\nu$ is smaller than the limiting thickness without light. As the frequency increases, the influence will be more and more important, until absorption of light in the oxide layer begins. It will perhaps be possible to observe a substantial effect for frequencies above the absorption band.

Inserting numerical values in G, we get

$$G \sim 10^{20} \text{ cm.}^2 \text{ sec./erg.}$$

Assuming a value $I \sim 10^5 \text{ erg./cm.}^2 \text{ sec.}$ for the intensity, we obtain $\log(G/I) \sim 35$, which can be erroneous by ± 5 , because of the rough estimation of different factors in G. For aluminium $\psi \sim 2 \text{ eV.}$, $\phi \sim 1 \text{ eV.}$, therefore

$$x_0 \sim 20 \text{ \AA}, \quad (23)$$

a quite reasonable value, because we know experimentally that for $h\nu \sim 4 \text{ eV.}$, the thickness at which the influence of light begins is $\sim 10 \text{ \AA}$, more or less the same as the theoretical value $x_0\psi/h\nu \sim 10 \text{ \AA}$. We should expect the influence of the light on the oxidation of aluminium at room temperature to begin for frequencies such that $h\nu > x_0\psi/x \sim 1 \text{ eV.}$

Let us study now the actual law of growth under the influence of light. The law of growth will be given as before by (5)

$$\frac{dx}{dt} = a\nu \exp \left\{ - \frac{U - e_a F}{kT} \right\}, \quad (24)$$

but the field F is now

$$F = (\psi + \xi)/ex.$$

For values of x bigger than $x_0\psi/h\nu$, we can use the expression (22) for ξ , and the field F becomes

$$F = (\psi + h\nu)/e(x + x_0). \quad (25)$$

For x smaller than $x_0\psi/h\nu$, F will be practically ψ/ex . It is interesting to see how this changes the limiting thickness x_c introduced by Mott. It is defined by the thickness for which the rate of growth is very small (say $\dot{x}_c = a \times 10^{-7} \text{ cm. sec.}^{-1}$); we deduce now from (24) the formula

$$\frac{x'_c + x_0}{a} = \frac{G'}{1 - (T/T_c)}; \quad G' = \frac{e_i \psi + h\nu}{e U}, \quad (26)$$

and T_c is given by (6). Putting the usual values for the constants appearing in (26), we obtain $x'_c \sim 60 \text{ \AA}$ at room temperature. This limiting thickness has not been reached in the experiments represented in fig. 5. Above $x_0\psi/h\nu \sim 10 \text{ \AA}$ we should expect a steady, almost linear, growth, as it is observed, because then the field (25) decreases very slowly.

§ 4. SUMMARY.

The theory of Mott (1940, 1947) for the oxidation of metals at low temperatures has been discussed and extended to oxides, such as Cu_2O , for which the metal diffuses through the oxide by the mechanism of vacant lattice points. It is also proved that the logarithmic law will be valid down to very low temperatures and for pressures of oxygen above $\sim 10^{-4}$ mm. Hg., independently of the temperature and the oxide considered. The model of Mott has also been applied to the explanation of the influence of light on the oxidation of aluminium, observed by Cabrera, Terrien and Hamon (1947). The theory agrees with the experimental results and suggests that the influence will be observable only for frequencies such that $h\nu > 2 \text{ eV.}$; the limiting thickness of the oxide layer increases by a factor $(\psi + h\nu)/\psi$ under the influence of light.

LITERATURE.

- CABRERA, N., 1945, *Comptes Rendus*, **220**, 111.
 CABRERA, N., and HAMON, J., 1947, *Comptes Rendus*, **224**, 1713.
 CABRERA, N., TERRIEN, J., and HAMON, J., 1947, *Comptes Rendus*, **224**, 1558.
 CAMPBELL, W. E., and THOMAS, U. B., 1947, *Trans. Electrochem. Soc.*, 345.
 GULBRANSEN, E. A., and WYSONG, W. S., 1947, *Jour. Phys. and Colloid Chem.*, **51**, 1087.
 GÜNTERSCHULTZE, A., and BETZ, H., 1934, *Zeits. f. Physik*, **92**, 367.
 LUSTMAN, B., and MEHL, R. F., 1941, *Trans. A.I.M.M.E.*, **143**, 246.
 MITCHELL, 1934, *Proc. Roy. Soc. A*, **146**, 442. See also WILSON, A. H., 1936, *Theory of Metals*, 142 (Cambridge).
 MOTT, N. F., 1940, *Trans. Far. Soc.*, **36**, 472; 1947, *Ibid.*, **43**, 429; 1947, *Journ. de Chim. Physique*, **44**, 172; MOTT and GURNEY, 1948, *Electronic Processes*, 267 (Oxford).
 VERWEY, E. J. W., 1935, *Physica*, **2**, 1059.
 WAGNER, C., 1933, *Zeits. f. Phys. chem.*, B **21**, 25. See also BARDEEN, J., BRATTAIN, W. H., and SHOCKLEY, W., 1946, *J. Chem. Phys.*, **14**, 714.

XV. *The Application of Mellin Transforms to the Summation of Slowly Convergent Series.*

By G. G. MACFARLANE, Dr.-Ing., B.Sc.,
 Telecommunications Research Establishment, Ministry of Supply, Malvern*.

[Received November 6, 1947.]

SUMMARY.

A powerful method of summing slowly convergent series is described and illustrated. It depends on properties of Mellin transforms. A table of over 70 Mellin Transforms is appended.

* Communicated by the Author,

1. INTRODUCTION.

A PROBLEM of frequent occurrence in applied mathematics is the summation of a slowly convergent series of terms. In many cases succeeding terms oscillate in sign and fall off in magnitude so slowly that evaluation of the sum from the series as it stands is impracticable. There are a number of well-known methods available for summing such a series. Perhaps the most widely used method is to convert the series into a contour integral and to try to evaluate the integral with the aid of Cauchy's theorem. This often leads to a new series, which is rapidly convergent. Another method makes use of Poisson's formula (Titchmarsh, p. 60, 1937), which relates Fourier cosine transforms. A further method, which seems to be much less widely known, makes use of Mellin transforms. In the experience of the author the method of Mellin transforms is very powerful and it is especially useful when the series is of the form

$$\sum_{n=0}^{\infty} f\{(n+a)^{\nu}\}.$$

Although the essential idea of the method is given in the literature (Titchmarsh, p. 63, 1937), it is thought that its development and application in applied mathematics is worth describing. This paper is therefore written with the object not only of describing the method, but also of collecting together a list of Mellin transforms and of useful cognate formulæ.

2. Mellin's Inversion formulæ are

$$F(s)=\int_0^{\infty} f(x) \cdot x^{s-1} dx, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and

$$f(x)=\frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} F(s) \cdot x^{-s} ds \quad \sigma_1<\sigma<\sigma_2. \quad . \quad . \quad . \quad . \quad (2)$$

The variable s is complex and its real part is σ . The limits σ_1 and σ_2 define a vertical strip in the complex s -plane in which the integral for $F(s)$ is convergent. Classes of function that satisfy these inversion formulæ are defined in Titchmarsh (1939, 1.29), and in Doetsch (1933, para. 8).

A number of useful rules for deriving Mellin transforms are easily deduced from (1) and (2). Thus the following are pairs

$$\begin{array}{ll} f(x), & F(s) \quad \sigma_1<\sigma<\sigma_2, \\ x^{\nu}f(x), & F(s+\nu) \quad \sigma_1-\nu<\sigma<\sigma_2-\nu, \quad . \quad . \quad . \quad . \quad (3) \end{array}$$

$$f(x^{\nu}), \quad \frac{1}{\nu} F\left(\frac{s}{\nu}\right) \quad \nu\sigma_1<\sigma<\nu\sigma_2, \quad . \quad . \quad . \quad . \quad (4)$$

$$f(ax), \quad a^{-s}F(s) \quad \sigma_1<\sigma<\sigma_2. \quad . \quad . \quad . \quad . \quad (5)$$

3. The Application to Summation of Series depends on (2). In it replace x by $n+a$ and we get

$$f(n+a) = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} F(s) \cdot (n+a)^{-s} ds \quad \sigma_1 < \sigma < \sigma_2.$$

Now sum over n and interchange integration and summation

$$\begin{aligned} \sum_{n=0}^{\infty} f(n+a) &= \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} F(s) \cdot \sum_{n=0}^{\infty} (n+a)^{-s} ds \\ &= \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} F(s) \cdot \zeta(s, a) ds \quad \sigma_1 < \sigma < \sigma_2. \quad \dots \quad (6) \end{aligned}$$

The contour is now distorted in such a way as to enclose the poles of the integrand so that Cauchy's theorem can be used. In carrying out this process a knowledge of certain properties of the generalised zeta function is necessary.

4. Properties of $\zeta(s, a)$ for $0 < a \leq 1$.

(i) The only singularity of $\zeta(s, a)$ is a simple pole with residue $+1$ at $s=1$.

$$\lim_{s \rightarrow 1} \left\{ \zeta(s, a) - \frac{1}{s-1} \right\} = -\frac{\Gamma'(a)}{\Gamma(a)}.$$

$$(ii) \quad \zeta(-m, a) = \frac{-1}{m+1} B_{m+1}(a), \quad m=0, 1, 2, \dots$$

where $B_n(a)$ is the Bernoulli polynomial of the n th degree (Davis 1933)

$$\zeta(-m, 1-a) = (-1)^{m+1} \zeta(-m, a), \quad m=0, 1, 2, \dots$$

$$(iii) \quad \zeta(0, a) = \frac{1}{2} - a.$$

(iv) When $\text{Re}(s) = \sigma > 1$,

$$\zeta(s, a) = \sum_{n=0}^{\infty} (a+n)^{-s},$$

$$\zeta(s, 1) = \zeta(s) \text{ the Zeta function of Riemann.}$$

$$\zeta(0) = -\frac{1}{2},$$

$$\zeta(-2m) = 0, \quad m=1, 2, 3, \dots$$

$$\zeta(-2m+1) = \frac{(-1)^m}{2m} B_m, \quad m=0, 1, 2, \dots$$

$$\zeta(2m) = \frac{2^{2m-1} \pi^{2m} B_m}{\Gamma(2m+1)}, \quad m=1, 2, 3, \dots$$

where B_m is the m th Bernoulli number (Davis 1933).

$$(v) \quad \operatorname{Res}_{z=-n} \{\Gamma(z)\} = \frac{(-)^n}{\Gamma(n+1)},$$

$$\lim_{s \rightarrow 1} \left\{ \zeta(s) - \frac{1}{s-1} \right\} = \gamma = 0.5772157.$$

$$(vi) \quad 2^{1-s} \Gamma(s) \zeta(s) \cos \left(\frac{1}{2} s \pi \right) = \pi^s \zeta(1-s).$$

Therefore $\Gamma(s) \zeta(s)$ has poles at $s = -2m+1$, $m=0, 1, 2, \dots$.

$$(vii) \quad \sum_{n=1}^{\infty} n^{-s} = \zeta(s),$$

$$\sum_{n=1}^{\infty} (-)^{n+1} n^{-s} = (1-2^{1-s}) \zeta(s),$$

$$\sum_{n=0}^{\infty} (-)^n (2n+1)^{-s} = \left\{ \zeta\left(s, \frac{1}{4}\right) - \zeta\left(s, \frac{3}{4}\right) \right\} 2^{-2s},$$

$$\sum_{n=0}^{\infty} (2n+1)^{-s} = (1-2^{-s}) \zeta(s),$$

$$\sum_{n=0}^{\infty} (-)^n (n+a)^{-s} = \left\{ \zeta\left(s, \frac{a}{2}\right) - \zeta\left(s, \frac{1+a}{2}\right) \right\} \cdot 2^{-s}.$$

$$(viii) \quad \Gamma(s) \Gamma(1-s) = \frac{\pi}{\sin(\pi s)}.$$

5. Examples.

We shall now illustrate the method by working out a few examples. The first example is a well-known series, the second is a very slowly convergent series of Bessel functions, and the third is a complicated series with a finite number of terms.

(i) Consider the series

$$S = \sum_{n=1}^{\infty} \frac{\cos(ny)}{n^2}.$$

From entry 34 of the accompanying table of Mellin transforms the transform of

$$f(x) = \cos x \text{ is } F(s) = \frac{2^{s-1} \sqrt{\pi} \Gamma(\frac{1}{2}s)}{\Gamma(\frac{1}{2} - \frac{1}{2}s)}, \quad 0 < \sigma < 1.$$

Therefore by rule (3)

$$\frac{\cos x}{x^2} \quad \text{and} \quad \frac{2^{s-3} \sqrt{\pi} \Gamma(\frac{1}{2}s-1)}{\Gamma(\frac{3}{2} - \frac{1}{2}s)}, \quad 2 < \sigma < 3,$$

are Mellin transforms. Hence

$$\sum_{n=1}^{\infty} \frac{\cos(ny)}{(ny)^2} = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \frac{2^{s-3} \sqrt{\pi} \Gamma(\frac{1}{2}s-1)}{\Gamma(\frac{3}{2}-\frac{1}{2}s)} y^{-s} \zeta(s) ds, \quad 2 < \sigma < 3.$$

But $\zeta(s)\Gamma(\frac{1}{2}s-1)$ has poles only at $s=2, 1, 0$ with residues

$$y^{-2}\zeta(2), \quad y^{-1}\frac{1}{2}\pi, \quad -\frac{1}{2}\zeta(0),$$

respectively, and these are the only poles of the integrand. Moreover, as $s \rightarrow -\infty$, the integrand tends to zero. Therefore, by Cauchy's theorem,

$$\sum_{n=1}^{\infty} \frac{\cos(ny)}{(ny)^2} = \frac{\pi^2}{6y^2} - \frac{\pi}{2y} + \frac{1}{4}.$$

(ii) The second series to be summed is

$$S = \sum_{n=0}^{\infty} \frac{(-)^n}{(2n+1)} J_1(\overline{(2n+1)}, y),$$

where J_1 is the Bessel function of the first order.

From entry 35 in the table and rules 3 and 5 we get the pair

$$f(x) = \frac{J_1(xy)}{x}, \quad F(s) = \frac{1}{2} \left(\frac{y}{2}\right)^{1-s} \frac{\Gamma(\frac{1}{2}s)}{\Gamma(2-\frac{1}{2}s)}, \quad 0 < \sigma < \frac{5}{2}.$$

Therefore, using the third formula of 4 (vii),

$$S = \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \frac{1}{2} \left(\frac{y}{2}\right)^{1-s} \frac{\Gamma(\frac{1}{2}s)}{\Gamma(2-\frac{1}{2}s)} \{\zeta(s, \frac{1}{4}) - \zeta(s, \frac{3}{4})\} 2^{-2s} ds, \quad 0 < \sigma < \frac{5}{2}.$$

The integrand has poles only at $s=-2n$ due to the factor $\Gamma(\frac{1}{2}s)$. Using Cauchy's theorem, we get

$$\begin{aligned} S &= \sum_{n=0}^{\infty} \frac{(-)^n (2y)^{2n+1}}{4\Gamma(n+1)\Gamma(n+2)} \{\zeta(-2n, \frac{1}{4}) - \zeta(-2n, \frac{3}{4})\} \\ &= - \sum_{n=0}^{\infty} \frac{1}{2} \frac{(-)^n B_{2n+1}(\frac{1}{4})(2y)^{2n+1}}{(2n+1)\Gamma(n+1)\Gamma(n+2)}, \end{aligned}$$

which is rapidly convergent for $y < 1$.

(iii) Consider the finite series

$$S = \sum_{m=1}^M \frac{(1-xm^{\frac{1}{3}})^{\frac{1}{3}}}{m^{\frac{1}{3}}},$$

where M is the largest integer such that $xM^{\frac{1}{3}} < 1$.

In this case we arrive at an asymptotic series, which provides an excellent approximation to the sum for small values of x , that is when the series as it stands has a large number of terms.

From entry 22 of the table we get the pair

$$f_1(x) = \begin{cases} (1-x)^{\frac{1}{2}} & 0 < x < 1 \\ 0 & x \geq 1 \end{cases} \quad F_1(s) = \frac{\Gamma(s)\Gamma(\frac{3}{2})}{\Gamma(s+\frac{3}{2})}, \quad \sigma > 0.$$

Using rule (4) with $\nu = \frac{2}{3}$ gives the pair

$$f_2(x) = \begin{cases} (1-x^{\frac{2}{3}})^{\frac{1}{2}} & 0 < x < 1 \\ 0 & x \geq 1 \end{cases} \quad F_2(s) = \frac{3}{2} \frac{\Gamma(\frac{3}{2}s)\Gamma(\frac{3}{2})}{\Gamma(\frac{3}{2}s+\frac{3}{2})}, \quad \sigma < 0.$$

Again, using rule (3) with $\nu = -\frac{2}{3}$ gives

$$f(x) = \begin{cases} \frac{(1-x^{\frac{2}{3}})^{\frac{1}{2}}}{x^{\frac{2}{3}}} & 0 < x < 1 \\ = 0 & x \geq 1. \end{cases} \quad F(s) = \frac{3}{2} \frac{\Gamma(\frac{3}{2}s-1)\Gamma(\frac{3}{2})}{\Gamma(\frac{3}{2}s+\frac{1}{2})}, \quad \sigma > \frac{2}{3}.$$

Hence

$$\begin{aligned} \sum_{m=1}^{\infty} f(xm) &= \sum_{m=1}^M \frac{(1-x^{\frac{2}{3}}m^{\frac{2}{3}})^{\frac{1}{2}}}{(xm)^{\frac{2}{3}}}, \\ &= \frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} \frac{3}{2} \frac{\Gamma(\frac{3}{2})\Gamma(\frac{3}{2}s-1)}{\Gamma(\frac{3}{2}s+\frac{1}{2})} x^{-s} \zeta(s) ds. \end{aligned}$$

$\Gamma(\frac{3}{2}s-1)$ has poles at $s = \frac{2}{3}, 0, -\frac{2}{3}, \dots, -(n-1)\frac{2}{3}$ with residues $\frac{2}{3} \frac{(-)^n}{\Gamma(n+1)}$; $\zeta(s)$ has a pole at $s=1$ with residue 1.

In this case, however, the sum of the residues at the poles gives only an asymptotic series because the contribution from the contour enclosing the poles at $s = -\infty$ is not zero, although it tends to zero as x tends to zero. The asymptotic series is

$$\sum_{m=1}^M \frac{\sqrt{1-xm^{\frac{2}{3}}}}{m^{\frac{2}{3}}} \sim \frac{3\pi}{4\sqrt{x}} + \zeta(\frac{2}{3}) + \frac{1}{4}x - \frac{1}{8}\zeta(-\frac{2}{3})x^2 + \dots$$

These three examples should suffice to show the power and elegance of the method. By ringing the changes on rules (3), (4) and (5), it is possible to deal with some very complicated series. An interesting identity that can be readily proved by this method is

$$\sum_{n=-\infty}^{+\infty} \frac{\sin^2(n\pi+\theta)}{(n\pi+\theta)^2} \equiv 1.$$

In conclusion we give a table of Mellin transforms collected from various sources. It does not attempt to be exhaustive, but it includes many useful examples.

TABLE OF MELLIN TRANSFORMS

	$f(x)$	$F(s)$	
1	$\frac{1}{2\pi i} \int_{\sigma-i\infty}^{\sigma+i\infty} F(s) x^{-s} ds$	$\int_0^\infty f(x) x^{s-1} dx$	$\sigma_1 < \sigma < \sigma_2$
2	$f(ax)$	$a^{-s} F(s)$	$\sigma_1 < \sigma < \sigma_2$
3	$x^v f(x)$	$F(s+v)$	$\sigma_1 - v < \sigma < \sigma_2 - v$
4	$f(x^v)$	$\frac{1}{v} F(\frac{s}{v})$	$v\sigma_1 < \sigma < v\sigma_2$
5	$f(\frac{1}{x})$	$F(-s)$	$-\sigma_2 < \sigma < -\sigma_1$
6	$\frac{d f(x)}{d_x x}$	$-(s-1) F(s-1)$	
7	$\int_0^x f(x) dx$	$-\frac{1}{s} F(s+1)$	$\sigma < 0$
8	$\int_x^\infty f(x) dx$	$\frac{1}{s} F(s+1)$	$\sigma > 0$
9	$f(x) \log(x)$	$\frac{d F(s)}{ds}$	
10	$f(x) g(x)$	$\frac{1}{2\pi i} \int_{K-i\infty}^{K+i\infty} F(w) G(s-w) dw$	
11	$\int_0^\infty g(u) f(\frac{x}{u}) \frac{du}{u}$	$F(s) G(s)$	
12	e^{-x}	$\Gamma(s)$	$\sigma > 0$
13	$\frac{1}{e^x - 1}$	$\Gamma(s) \zeta(s)$	$\sigma > 1$
14	$\frac{1}{e^x + e^{-x}}$	$\Gamma(s) L(s)$	$\sigma > 0$
15	$e^{-x} \log x$	$\Gamma'(s)$	$\sigma > 0$
16	$\exp(-x^2)$	$\frac{1}{2} \Gamma(\frac{1}{2}s)$	$\sigma > 0$
17	$\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) dt$	$\frac{1}{\sqrt{\pi} s} \Gamma(\frac{1}{2}s + \frac{1}{2})$	$\sigma > 0$
18	$\frac{1}{1+x}$	$\pi \operatorname{cosec}(\pi s)$	$0 < \sigma < 1$

19	$\frac{1}{(1+x)^v}$	$\frac{\Gamma(s)\Gamma(v-s)}{\Gamma(v)}$	$0 < \sigma < R(v)$
20	$\frac{1}{x^2+1}$	$\frac{\pi}{2} \coth \left(\frac{\pi s}{2} \right)$	$0 < \sigma < 2$
21	$\begin{matrix} 1 & x < a \\ 0 & x \geq a \end{matrix}$	$\frac{a^s}{s}$	$\sigma > 0$
22	$\begin{matrix} (1-x)^{v-1} & 0 < x < 1 \\ 0 & x \geq 1 \end{matrix}$	$\frac{\Gamma(s)\Gamma(v)}{\Gamma(s+v)}$	$\sigma > 0$ $R(v) > 0$
23	$\begin{matrix} 0 & 0 < x \leq 1 \\ (x-1)^{-v} & x > 1 \end{matrix}$	$\frac{\Gamma(v-s)\Gamma(1-v)}{\Gamma(1-s)}$	$\sigma < R(v) < 1$
24	$\begin{matrix} (x^2-1)^{-v-\frac{1}{2}} & x > 1 \\ 0 & 0 < x \leq 1 \end{matrix}$	$\frac{\Gamma(v+\frac{1}{2}-\frac{1}{2}s)\Gamma(\frac{1}{2}-v)}{2\Gamma(1-\frac{1}{2}s)}$	$\sigma < 2R(v+\frac{1}{2}) < 2$
25	$\begin{matrix} \sqrt{1-x^2} & 0 < x \leq 1 \\ 0 & x > 1 \end{matrix}$	$\frac{1}{2} \frac{\Gamma(\frac{3}{2}s)\Gamma(\frac{3}{2})}{\Gamma(\frac{3}{2}s+\frac{3}{2})}$	$\sigma > 0$
26	$\begin{matrix} 0 & 0 < x \leq 1 \\ \frac{(x-\sqrt{x^2-1})^v + \{x-\sqrt{x^2-1}\}^v}{\sqrt{x^2-1}} & x > 1 \end{matrix}$	$\frac{2^{-s}\Gamma(\frac{1}{2}v+\frac{1}{2}-\frac{1}{2}s)\Gamma(\frac{1}{2}-\frac{1}{2}s-\frac{1}{2}v)}{\Gamma(1-s)}$	$\sigma < R(v) + 1$
27	$\frac{\{\sqrt{x^2+1}-x\}^v}{\sqrt{x^2+1}}$	$\frac{2^{-s}\Gamma(s)\Gamma(\frac{1}{2}+\frac{1}{2}v-\frac{1}{2}s)}{\Gamma(\frac{1}{2}+\frac{1}{2}v+\frac{1}{2}s)}$	$0 < \sigma < R(v) + 1$
28	$\log(1+x)$	$\frac{\pi}{s \sin(\pi s)}$	$\sigma < 0$
29	$\begin{matrix} \log(\frac{a}{x}) & x < a \\ 0 & x \geq a \end{matrix}$	$\frac{a^s}{s^2}$	$\sigma > 0$
30	$\log \left \frac{1+x}{1-x} \right $	$\frac{\pi}{s} \tan \left(\frac{\pi s}{2} \right)$	$-1 < \sigma < 1$
31	$-\log \left(2 \sin \left \frac{x}{2} \right \right)$	$\Gamma(s) \cos \left(\frac{\pi s}{2} \right) \zeta(s+1)$	$0 < \sigma < 1$
32	$F(a, b; c; -x)$	$\frac{\Gamma(s)\Gamma(a-s)\Gamma(b-s)}{\Gamma(c-s)} \frac{\Gamma(c)}{\Gamma(a)\Gamma(b)}$	$0 < \sigma < \min\{R(a), R(b)\}$
33	$\sin(x)$	$\frac{2^{s-1}\sqrt{\pi}\Gamma(\frac{1}{2}+\frac{1}{2}s)}{\Gamma(1-\frac{1}{2}s)} = \Gamma(s) \sin \left(\frac{\pi s}{2} \right)$	$-1 < \sigma < 1$
34	$\cos(x)$	$\frac{2^{s-1}\sqrt{\pi}\Gamma(\frac{1}{2}s)}{\Gamma(\frac{1}{2}-\frac{1}{2}s)} = \Gamma(s) \cos \left(\frac{\pi s}{2} \right)$	$0 < \sigma < 1$
35	$J_v(x)$	$\frac{2^{s-1}\Gamma(\frac{1}{2}s+\frac{1}{2}v)}{\Gamma(\frac{1}{2}v-\frac{1}{2}s+1)}$	$-v < \sigma < \frac{3}{2}$
36	$Y_v(x)$	$-\frac{2^{s-1}}{\pi} \Gamma(\frac{1}{2}s+\frac{1}{2}v)\Gamma(\frac{1}{2}s-\frac{1}{2}v) \cos(s-v) \frac{\pi}{2}$	$ v < \sigma < \frac{3}{2}$
37	$J_v(x) + J_{-v}(x)$	$\frac{2^s}{\pi} \Gamma(\frac{1}{2}s+\frac{1}{2}v)\Gamma(\frac{1}{2}s-\frac{1}{2}v) \sin \left(\frac{\pi s}{2} \right) \cos \left(\frac{\pi v}{2} \right)$	$ v < \sigma < \frac{3}{2}$
38	$J_v(x) - J_{-v}(x)$	$-\frac{2^s}{\pi} \Gamma(\frac{1}{2}s+\frac{1}{2}v)\Gamma(\frac{1}{2}s-\frac{1}{2}v) \cos \left(\frac{\pi s}{2} \right) \sin \left(\frac{\pi v}{2} \right)$	$ v < \sigma < \frac{3}{2}$

39	$\cos(x) J_\nu(x)$	$\frac{2^{s-1} \sqrt{\pi} \Gamma(\frac{1}{2}s + \frac{1}{2}\nu) \Gamma(\frac{1}{2}s)}{\Gamma(1 + \frac{1}{2}\nu - \frac{1}{2}s) \Gamma(\frac{1}{2} - \frac{1}{2}\nu - \frac{1}{2}s) \Gamma(\frac{1}{2} + \frac{1}{2}\nu - \frac{1}{2}s)}$	$-R(\nu) < \sigma < \frac{1}{2}$
40	$\sin(x) J_\nu(x)$	$\frac{2^{s-1} \sqrt{\pi} \Gamma(\frac{1}{2}s + \frac{1}{2}\nu + \frac{1}{2}) \Gamma(\frac{1}{2}s)}{\Gamma(1 + \frac{1}{2}\nu - \frac{1}{2}s) \Gamma(1 - \frac{1}{2}\nu - \frac{1}{2}s) \Gamma(\frac{1}{2} + \frac{1}{2}\nu - \frac{1}{2}s)}$	$-R(\nu+1) < \sigma < \frac{1}{2}$
41	$e^{ix} J_\nu(x)$	$\frac{2^{-s} \Gamma(s+\nu) \Gamma(\frac{1}{2}s)}{\sqrt{\pi} \Gamma(1+\nu-s)} e^{i(\frac{1}{2}+\nu)\frac{\pi}{2}}$	$-R(\nu) < \sigma < \frac{1}{2}$
42	$H_\nu^{(0)}(x)$	$-\frac{i2^{s-1}}{\pi} \Gamma(\frac{1}{2}s + \frac{1}{2}\nu) \Gamma(\frac{1}{2}s - \frac{1}{2}\nu) e^{i(\frac{1}{2}-\nu)\frac{\pi}{2}}$	$ V < \sigma < \frac{3}{2}$
43	$K_\nu(x)$	$2^{s-2} \Gamma(\frac{1}{2}s - \frac{1}{2}\nu) \Gamma(\frac{1}{2}s + \frac{1}{2}\nu)$	$\sigma + \nu > \max(0, 2\nu)$
44	$H_\nu(x)$	$\frac{2^{s-1} \Gamma(\frac{1}{2}s + \frac{1}{2}\nu) \tan(s+\nu)\frac{\pi}{2}}{\Gamma(\frac{1}{2}\nu - \frac{1}{2}s + 1)}$	$-1 < \sigma + \nu < \nu + \frac{3}{2}$
45	$e^{-x} I_\nu(x)$	$\frac{\Gamma(s+\nu) \Gamma(\frac{1}{2}s)}{2^s \sqrt{\pi} \Gamma(1+\nu-s)}$	
46	$e^x K_\nu(x)$	$\frac{2^{-s}}{\sqrt{\pi}} \cos(\nu\pi) \Gamma(\frac{1}{2}s-s) \Gamma(s+\nu) \Gamma(s-\nu)$	
47	$\frac{2}{\pi} K_0(x) - Y_0(x)$	$\frac{1}{\pi} \Gamma^2(\frac{1}{2}s) \cos^2(\frac{1}{4}s\pi) 2^s$	$\sigma > 1$
48	$J_\nu(ax) \exp(-\frac{1}{2}x^2)$	$\frac{\Gamma(\frac{1}{2}\nu + \frac{1}{2}s) (\frac{a}{2b})^\nu}{2b^s \Gamma(s+1)} {}_2F_1(\frac{1}{2}\nu + \frac{1}{2}s; \nu+1; -\frac{a^2}{4b^2})$	$R(s+\nu) > 0$
49	$J_\nu(bx) e^{-ax}$	$\frac{\Gamma(s+\nu) (\frac{b}{2a})^\nu}{a^s \Gamma(\nu+1)} {}_2F_1(\frac{s+\nu}{2}, \frac{s+\nu+1}{2}; \nu+1; -\frac{b^2}{a^2})$	$R(a+ib) > 0$ $R(a-ib) > 0$
50	$\frac{J_\mu(x) J_\nu(x)}{x^{\mu+\nu}}$	$\frac{2^{s-\mu-\nu-1} \Gamma(\frac{1}{2}s) \Gamma(1+\mu+\nu-s)}{\Gamma(1+\nu-\frac{1}{2}s) \Gamma(1+\mu-\frac{1}{2}s) \Gamma(1+\mu+\nu-\frac{1}{2}s)}$	$0 < \sigma < R(\mu+\nu+1)$
51	$\frac{K_\mu(x) K_\nu(x)}{x^{\mu+\nu}}$	$\frac{2^{s-\mu-\nu-1} \Gamma(\frac{1}{2}s) \Gamma(\frac{1}{2}s-\mu) \Gamma(\frac{1}{2}s-\nu) \Gamma(\frac{1}{2}s-\mu-\nu)}{\Gamma(s-\mu-\nu)}$	
52	$x^{-2\nu} J_\nu(x) K_\nu(x)$	$\frac{2^{s-2\nu-2} \Gamma(\frac{1}{4}s) \Gamma(\frac{1}{4}s-\nu)}{\Gamma(1+\nu-\frac{1}{4}s)}$	
53	$J_\nu(x) Y_\nu(x)$	$\frac{-1}{2\sqrt{\pi}} \frac{\Gamma(\frac{1}{2}s) \Gamma(\frac{1}{2}s+\nu)}{\Gamma(\frac{1}{2}s+\frac{1}{2}) \Gamma(1+\nu-\frac{1}{2}s)}$	
54	$I_\nu(x) K_\nu(x)$	$\frac{1}{4\sqrt{\pi}} \frac{\Gamma(\frac{1}{2}s+\nu) \Gamma(\frac{1}{2}s) \Gamma(\frac{1}{2}-\frac{1}{2}s)}{\Gamma(1+\nu-\frac{1}{2}s)}$	
55	$e^{-x \cosh d} I_\nu(x \sinh d)$	$\Gamma(s+\nu) P_{s-1}^{-\nu}(\cosh d)$	$R(s+\nu) > 0$
56	$e^{-x \cosh d} K_\nu(x \sinh d)$	$\frac{\sin(s-1)\pi}{\sin(s+\nu-1)\pi} \Gamma(s-\nu) Q_{s-1}^\nu(\cosh d)$	$R(s) > R(\nu) $
57	$e^{-x \cos \beta} J_\nu(x \sin \beta)$	$\Gamma(s+\nu) P_{s-1}^{-\nu}(\cos \beta)$	$R(s+\nu) > 0$
58	$e^{-x \cos \beta} Y_\nu(x \sin \beta)$	$-\frac{\sin(s-1)\pi}{\sin(s+\nu-1)\pi} \frac{\Gamma(s-\nu)}{\pi} \times [Q_{s-1}^\nu(\cos \beta + 0i) e^{\frac{1}{2}\nu\pi i} + Q_{s-1}^\nu(\cos \beta - 0i) e^{-\frac{1}{2}\nu\pi i}]$	$R(s+\nu) > 0$

59	$e^{-x \cosh \alpha} I_\nu(x)$	$\frac{\cos(v\pi)}{\sin(s+v)\pi} \frac{Q_{\nu-\frac{1}{2}}^{s-\frac{1}{2}}(\cosh \alpha)}{\sqrt{v/2} \sinh^{s-\frac{1}{2}} \alpha}$	$R(s+v) > 0$ $R(\cosh \alpha) > 1$
60	$e^{-x \cosh \alpha} K_\nu(x)$	$\sqrt{\frac{1}{2}\pi} \Gamma(s-v) \Gamma(s+v) \frac{P_{\nu-\frac{1}{2}}^{s-\frac{1}{2}}(\cosh \alpha)}{\sinh^{s-\frac{1}{2}} \alpha}$	$R(\mu) > R(\nu) $ $R(\cosh \alpha) > -1$
61	$\frac{J_\nu \{a \sqrt{(x^2+z^2)}\}}{(x^2+z^2)^{\frac{1}{2}\nu}}$	$\frac{2^{\frac{1}{2}s-1} \Gamma(\frac{1}{2}s)}{a^{\frac{1}{2}s} z^{v-\frac{1}{2}s}} J_{\nu-\frac{1}{2}s}(az)$	$a \geq 0$ $2 \angle \sigma \angle R(v+\frac{1}{2})$
62	$\frac{K_\nu \{a \sqrt{(x^2+z^2)}\}}{(x^2+z^2)^{\frac{1}{2}\nu}}$	$\frac{2^{\frac{1}{2}s-1} \Gamma(\frac{1}{2}s)}{a^{\frac{1}{2}s} z^{v-\frac{1}{2}s}} K_{\nu-\frac{1}{2}s}(az)$	$a > 0$ $\sigma > 0$
63	$\frac{1}{(1+x)^m} P_{m-1} \left(\frac{1-x}{1+x} \right)$	$\frac{\Gamma(s) \{\Gamma(m-s)\}^2}{\Gamma(1-s) \{\Gamma(m)\}^2}$	$0 \angle \sigma \angle m$
64	$e^{\frac{1}{2}x} W_{k,m}(x)$	$\frac{\Gamma(s-k) \Gamma(s-m+\frac{1}{2}) \Gamma(s+m+\frac{1}{2})}{\Gamma(s-k-m+\frac{1}{2}) \Gamma(-k+m+\frac{1}{2})}$	
65	$\exp(\frac{1}{4}x^2) D_n(x)$	$\frac{\Gamma(-\frac{1}{2}s-\frac{1}{2}n) \Gamma(s)}{\Gamma(-n)} 2^{\frac{1}{2}(s-n-2)}$	
66	$\begin{matrix} \Xi^{-x} & 0 \angle x \angle a \\ 0 & a \angle x \end{matrix}$	$\Gamma(s, a)$	
67	$Ci(x)$	$\frac{\Gamma(s)}{s} \cos\left(\frac{\pi s}{2}\right)$	$1 \angle \sigma \angle 2$
68	$Si(x) - \frac{\pi}{2}$	$\frac{\Gamma(s)}{s} \sin\left(\frac{\pi s}{2}\right)$	$0 \angle \sigma \angle 2$
69	$-Ei(-x)$	$\frac{\Gamma(s)}{s}$	$\sigma > 1$
70	$\frac{\pi}{2} - \tan^{-1}(ax)$	$\frac{\pi}{2s} \sec\left(\frac{\pi s}{2}\right)$	$0 \angle \sigma \angle 2$
71	$\int_0^x \frac{x^\nu dx}{e^x - 1}$	$-\frac{1}{s} \Gamma(s+v+1) \zeta(s+v+1)$	$1-v \angle \sigma \angle 0$
72	$Ai(x)$	$\frac{3^{\frac{1}{3}s-\frac{2}{3}}}{2\pi} \Gamma(\frac{1}{3}s) \Gamma(\frac{1}{3}s+\frac{1}{3})$	$\sigma > 0$

ACKNOWLEDGEMENT.

The author desires to thank the Ministry of Supply for permission to publish this paper, which is Crown Copyright reserved and reproduced with the permission of the Controller of H.M. Stationery Office.

REFERENCES.

- DAVIS, M. T., 1933, *Tables of the Higher Mathematical Functions*, The Principia Press, Inc.
 DOETSCH, G., 1937, *Theorie und Anwendung der Laplace-Transformation*, Springer.
 TITCHMARSH, E. C., 1939, *Introduction to the Theory of Fourier Integrals*, OUP.

XVI. *The Brillouin Zones for the Co_2Al_3 and NiAl_3 Structures.*

By G. V. RAYNOR and M. B. WALDRON,
Department of Metallurgy, The University, Edgbaston*.

[Received October 6, 1948.]

1. INTRODUCTION.

IN recent years the constitutions of a number of aluminium-rich ternary systems, containing as solutes one or two transitional metals of the first long period, have been examined. This work, which has been described in several recent papers (Raynor 1944, 1945; Raynor and Little 1945; Raynor and Pfeil 1946-47; Raynor and Wakeman 1947; Raynor and Waldron 1948), was undertaken in order to obtain information with regard to the rôle played by transitional metals in this type of alloy. The results, taken together, strongly support the view that, when alloyed with a metal of relatively high valency such as aluminium, transitional metal atoms absorb electrons from the structure as a whole in such a manner as to fill the vacancies in their atomic orbitals. The importance of the electron : atom ratio in determining the formation of binary and ternary intermediate phases with aluminium and the transitional metals has also been demonstrated.

According to the theory of the electronic structure of transitional metals put forward by Pauling (1938), electrons derived from both the $3d$ and $4s$ levels of the free atom are to be regarded as concerned in cohesion in the solid state. For chromium, manganese, iron, cobalt and nickel, 5.78 electrons per atom occupy "bonding orbitals", and are concerned only with cohesion, while the remainder occupy "atomic orbitals", which can contain a maximum of 4.88 electrons per atom. Thus, chromium, manganese and iron have respectively 0.22, 1.22 and 2.22 electrons per atom in atomic orbitals; in each atom these electrons are "unpaired", and have the same value of the spin quantum number. In the case of iron, the 2.22 unpaired atomic orbital electrons account for the saturation magnetic moment, at absolute zero, of 2.22 Bohr magnetons per atom. On passing to cobalt, the addition of a further electron leads to the pairing of a certain number of atomic orbital electrons with electrons of opposite spin. As a consequence of this, cobalt has 1.51 paired electrons and 1.71 unpaired electrons per atom in atomic orbitals. Nickel, with one electron per atom more than cobalt, has 3.61 paired electrons and 0.61 unpaired electrons per atom in these orbitals.

* Communicated by the Authors.

According to this conception, there are vacancies in the atomic orbitals of the transitional metals of the first long period, to the following extents per atom :

Chromium	Manganese	Iron	Cobalt	Nickel
4.66	3.66	2.66	1.71	0.61

If it be accepted that, in aluminium-rich alloys, these vacancies may be filled by the absorption of electrons from the structure as a whole, several interesting analogies between different alloy systems may be traced. Thus, the systems aluminium-manganese-copper (Raynor 1944), aluminium-manganese-nickel (Raynor 1944) and aluminium-manganese-zinc (Raynor and Wakeman 1947) each contain a ternary compound which enters into equilibrium with the aluminium-rich primary solid solution; in each case the ternary compound is based upon a ratio of four aluminium atoms to one solute atom, and, if manganese and nickel are considered to absorb respectively 3.66 and 0.61 electrons per atom, has an electron : atom ratio of 1.85. It has also been shown that the binary compound Co_2Al_9 and the ternary compound FeNiAl_9 , the isomorphy of which may be accounted for by the fact that both phases have closely similar electron : atom ratios according to the present theory, will dissolve nickel to the same electron : atom ratio of 2.285 (Raynor and Pfeil 1946-47). On the basis of the theory, almost quantitative predictions of the forms of the equilibrium diagrams for the aluminium-rich aluminium-iron-cobalt (Raynor and Waldron 1948) and aluminium-cobalt-nickel alloys (Raynor and Pfeil 1946-47) have been made.

The view that absorption of electrons occurs when transitional metals are alloyed with aluminium has received strong support from the work of Mrs. Douglas (1948), who has determined the crystal structure of the intermediate phase Co_2Al_9 and has estimated the number of electrons contained within the cobalt peaks in the (hOl) and (hkO) Fourier syntheses. The results indicate that in Co_2Al_9 each cobalt atom has of the order of two electrons more than its normal number of 27.

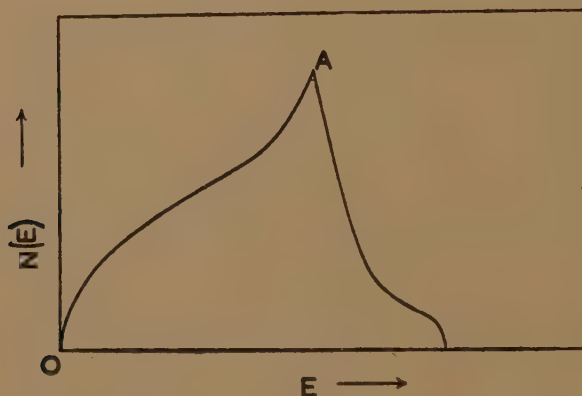
In the present paper, the Brillouin Zone characteristics of Co_2Al_9 , and of NiAl_3 , are examined, since, if the formation of such compounds is conditioned by electronic factors as suggested by the results described above, the forms of the Brillouin Zones would be expected to bear some relation to the observed electron : atom ratios.

2. THE BRILLOUIN ZONE STRUCTURE FOR Co_2Al_9 .

According to the work of Mrs. Douglas, Co_2Al_9 has a monoclinic structure with $a=8.5575$ Å, $b=6.290$ Å and $c=6.2130$ Å; the angle β is 94.760° . Each cobalt atom has a regular array of nine nearest neighbours. The Co-Al distances lie between 2.38 and 2.52 Å, while the Al-Al distances are within the range 2.70 to 3.00 Å. There are 22 atoms per unit cell, and the mean volume per atom is 15.149 Å³.

In the case of the normal type of electron compounds which occur in the alloys of copper, silver, and gold, the Brillouin Zones tend to be filled to a level which corresponds with slightly more electrons per atom than could be contained in the volume of the sphere in k -space which just touches the bounding planes of the zone (the so-called inscribed sphere). Thus the inscribed sphere for the face-centred cubic structure contains 1.36 electrons per atom, whereas the primary solid solubility limit occurs at 1.4 electrons per atom. Similarly, the body-centred cubic structures (β phases), which occur at 1.5 electrons per atom, are characterized by a Brillouin Zone with an inscribed sphere which can contain 1.48 electrons per atom. In a general way, it can be said that electron compounds tend to occur at compositions close to those which correspond to the number of electrons per atom contained in the inscribed sphere to the appropriate Brillouin Zone. This may be understood by

Fig. 1.

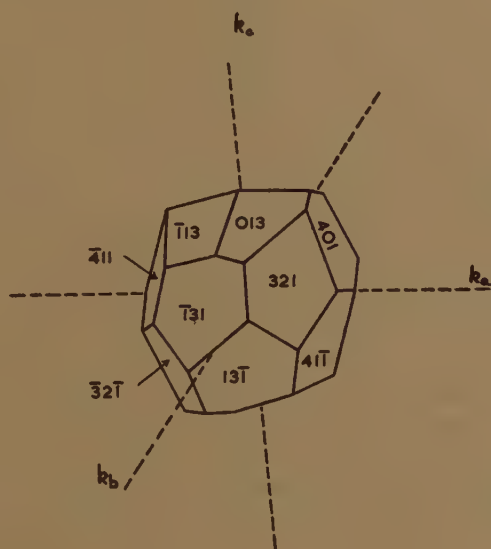


reference to the $N(E)$ curve for a Brillouin Zone, which expresses the electron distribution in terms of the energy (fig. 1). The point A represents the stage at which the number of electrons per atom present just fills the inscribed sphere. Any increase of the number of electrons per atom above this level will lead to a sudden increase in the energy of the structure, owing to the falling $N(E)$ curve. Structures will, however, tend to be formed such that the corresponding zone is filled with electrons to approximately the point A, since by this means the energy of the structure is relatively depressed. The exact electron:atom ratios at which such electron compounds are stable depend on the details of the $N(E)$ curves for the electron compound and the phase with which it is in equilibrium.

If, therefore, the phase Co_2Al_9 depends for its formation on the establishment of a certain electron:atom ratio, it would be expected that the inscribed sphere to the appropriate Brillouin Zone would contain a number of electrons per atom close to that calculated for Co_2Al_9 . According to the results of an investigation of primary crystals of the phase

extracted from slowly cooled alloys by anodic solution of the matrix (Raynor 1946-47) it contains slightly more cobalt than corresponds to the formula Co_2Al_9 . The electron:atom ratio for the experimentally determined composition (33.2 per cent Co) is 2.125, if each aluminium atom contributes three electrons while each cobalt atom absorbs 1.71 electrons. The sphere in k -space corresponding to this number of electrons per atom would have a radius of $\sqrt[3]{\frac{3}{4\pi} \cdot \frac{2.125}{2V}}$ where V is the volume per atom, or 0.2558 recip. Å. If this sphere were to touch the bounding planes of the appropriate Brillouin Zone, the planes with which contact was made would be at a perpendicular distance of $p=0.2558$

Fig. 2.



recip. Å from the origin of k -space. For Co_2Al_9 to be characterized by a Brillouin Zone whose inscribed sphere contains 2.125 electrons per atom, strong lines on the powder diffraction pattern should appear in such positions that the corresponding interplanar spacing of the diffracting planes was in the region of $\frac{1}{2p}$ or 1.9543 Å. For cobalt $K\alpha$ radiation, strong lines would thus be expected in the region of $\sin^2 \theta = 0.21$, where θ is the Bragg angle.

The diffraction pattern of Co_2Al_9 is complex, but contains two very strong lines, which occur, for cobalt $K\alpha$ radiation, at $\sin^2 \theta = 0.2090$ and 0.2131. The agreement of these values with that required for an inscribed sphere containing 2.125 electrons per atom is striking, and it is very probable that the planes giving rise to these reflections are also concerned in defining the form of the Brillouin Zone in k -space.

The two very strong lines are composed of overlapping reflections from the (401), ($\bar{4}$ 11), (013), (321), ($\bar{1}$ 31) and ($\bar{1}$ 13) planes. These planes form a closed volume in k -space of the form shown in fig. 2; the total volume corresponds approximately to 2.5 electrons per atom. The perpendicular distances of all the bounding planes of the zone from the origin are closely similar, and have the following values:

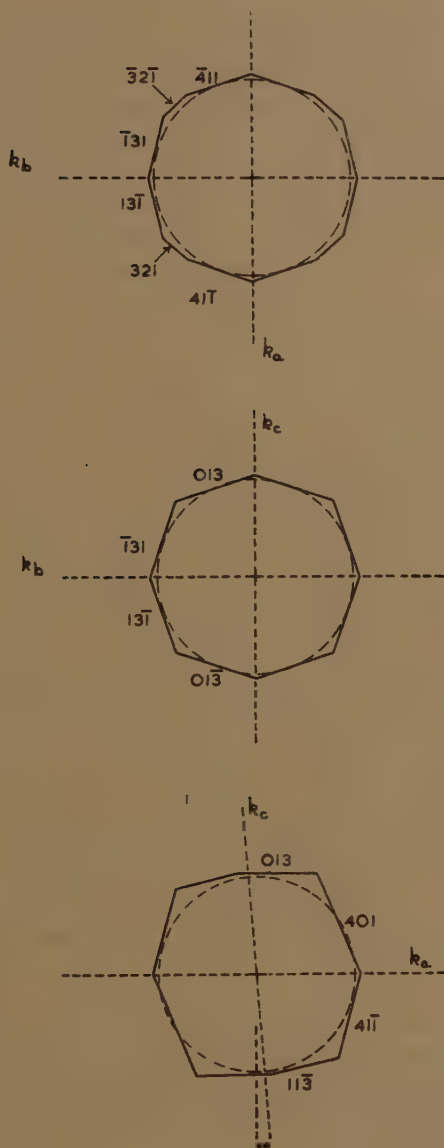
Plane	Perpendicular distance from origin of k -space
(401)	0.2543 recip. A.
($\bar{4}$ 11)	0.2544 " "
(013)	0.2550 " "
(321)	0.2551 " "
($\bar{1}$ 31)	0.2570 " "
($\bar{1}$ 13)	0.2571 " "

When the Fermi Sphere just touches the (401) plane, its volume corresponds to 2.1 electrons per atom; the volume of a sphere with radius equal to the perpendicular distance of the ($\bar{1}$ 13) plane from the origin corresponds with 2.15 electrons per atom. These values are very close to the value of 2.125 electrons per atom which is characteristic of the experimentally determined composition of Co_2Al_9 , and the probable Brillouin Zone structure is therefore in agreement with the theory of absorption of electrons by transitional metals in this type of compound.

It is of interest to consider more fully the form of the Brillouin Zone shown in fig. 2. Fig. 3 shows the sections in the $k_a k_b$, $k_b k_c$, and $k_a k_c$ planes. It will be seen that, in spite of the fact that the zone is bounded by six different planes, it is remarkably symmetrical in nature. The section in the $k_a k_b$ plane is almost circular, while the section in the $k_b k_c$ plane is also not far from circular. The section in the $k_a k_c$ plane is less symmetrical than the other two sections, but, by comparing figs. 2 and 3, it may be seen that the whole volume enclosed is approximately spherical.

The phase Co_2Al_9 therefore occurs at such a composition that the inscribed sphere to the appropriate Brillouin Zone contains approximately 2.12 electrons per atom. At this stage the zone restrictions on the direction of motion of the electrons first come into operation. The number of electrons per atom present can, however, be increased, since the solubility of nickel in both Co_2Al_9 and FeNiAl_9 proceeds until in each case an electron : ratio of 2.285 is reached. It is of interest to note that 2.285 electrons per atom corresponds with a sphere, in k -space, of radius 0.2601 recip. A. In fig. 3 the dotted circles represent sections of such a sphere, and it will be seen that the circles just touch the planes of the ($\bar{4}$ 11) type in the equatorial plane ($k_a k_b$ plane) and just touch the (013) and ($\bar{1}$ 31) planes in the $k_b k_c$ section. It appears, therefore, that only after this stage has been reached do the zone restrictions on the directions of motion and the energies of the electrons become sufficiently serious to give rise to the sudden increase in the energy of the structure which leads to instability.

Fig. 3.

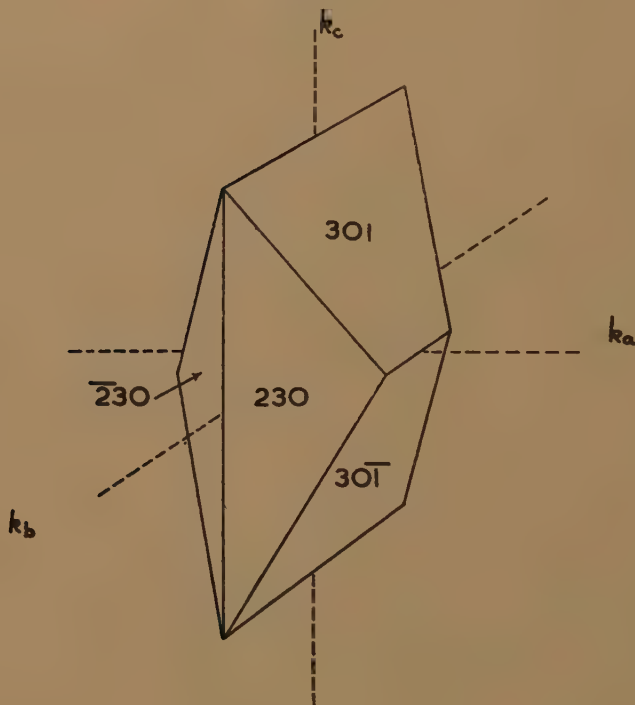


III. THE BRILLOUIN ZONE STRUCTURE FOR NiAl_3 .

The crystal structure of NiAl_3 has been determined by Bradley and Taylor (1937), and is orthorhombic, with $a=6.5982$ Å, $b=7.3515$ Å, and $c=4.8021$ Å. There are 16 atoms per unit cell, and the mean volume per atom is 14.5584 Å³. If the nickel atoms act as absorbers of electrons to the extent of 0.61 electrons per atom, the electron : atom ratio characteristic of NiAl_3 is 2.09 electrons per atom. Thus, if NiAl_3 , like Co_2Al_9 , is characterized by a Brillouin Zone capable of containing an inscribed

sphere corresponding to this number of electrons, strong lines should appear on the powder diffraction pattern in positions corresponding to an interplanar spacing of approximately 1.94 Å. For a photograph with cobalt $K\alpha$ radiation, this would correspond with strong lines in the region of $\sin^2 \theta = 0.212$. According to the work of Bradley and Taylor, the most intense lines in the diffraction pattern obtained with cobalt radiation (up to $\sin^2 \theta = 0.26$) are those due to reflection from the (301) and (230) planes. These reflections occur respectively at $\sin^2 \theta = 0.1996$ and 0.2062. The agreement with the expected value of 0.212 is good.

Fig. 4.



The Brillouin Zone defined in k -space by the planes (301) and (230) is shown in fig. 4. The perpendicular distances of these two planes from the origin of k -space are as follows:—

Plane	Perpendicular distance from origin of k -space.
(301)	0.2500 recip. Å.
(230)	0.2542 " "

The inscribed sphere to this Brillouin Zone, which just touches the (301) planes, but has not yet touched the (230) planes, corresponds to 1.91 electrons per atom. The Fermi sphere of radius equal to the perpendicular distance of the (230) planes from the origin corresponds to 2.00 electrons per atom. The Brillouin Zone for the NiAl_3 structure may therefore

be regarded as able to accommodate of the order of 2 electrons per atom before the zone restrictions on the directions of motion of the electrons cause the addition of further electrons to raise the energy of the structure rapidly. The agreement with the experimental electron:atom ratio of 2.09 is less exact than in the case of Co_2Al_9 , but is sufficiently close to confirm that NiAl_3 is to be considered as a type of electron compound, and that absorption of electrons by the nickel atoms takes place.

IV. SUMMARY.

The results described above indicate that, for the intermediate phases Co_2Al_9 and NiAl_3 , the most probable Brillouin Zones are characterized by inscribed spheres which can contain respectively 2.12 and 2.00 electrons per atom. These values are in close agreement with the values of 2.125 and 2.09 electrons per atom calculated for these structures according to the theory of the rôle of transitional metal solutes in aluminium-rich alloys developed in previous work. This work, combined with the observation of Mrs. Douglas that in Co_2Al_9 the cobalt atoms possess of the order of 2 electrons in excess of their normal complement of 27, confirms that, in aluminium-rich alloys, transitional metal atoms absorb electrons from the structure as a whole. The close quantitative agreement between the numbers of electrons per atom which correspond to the inscribed spheres of the Brillouin Zones and the numbers of electrons per atom calculated according to the absorption theory indicates that the number of electrons absorbed per atom of a transitional metal cannot differ appreciably from the number of vacancies per atom in its atomic orbitals.

The results described in this paper also give strong support to the view that the intermetallic binary and ternary compounds, formed by aluminium and transitional metals of the first long period, are properly to be considered as electron compounds rather than as compounds controlled mainly by electrochemical or atomic size considerations.

REFERENCES.

- BRADLEY, A. J., and TAYLOR, A., 1937 *Phil. Mag.*, **23**, 1049.
DOUGLAS, A. M. B., 1948, *Nature, Lond.*, **162**, 566.
PAULING, L., 1938, *Phys. Rev.*, (ii), **54**, 899.
RAYNOR, G. V., 1944, *J. Inst. Metals*, **70**, 531; 1945, *Phil. Mag.*, **36**, 770; 1946-47, *J. Inst. Metals*, **73**, 521.
RAYNOR, G. V., and LITTLE, K., 1945, *J. Inst. Metals*, **71**, 493.
RAYNOR, G. V., and PFEIL, P. C. L., 1946-47, *J. Inst. Metals*, **73**, 387, 609.
RAYNOR, G. V., and WAKEMAN, D. W., 1947, *Proc. Roy. Soc. A*, **190**, 82.
RAYNOR, G. V., and WALDRON, M. B., 1948, *Proc. Roy. Soc. A*, **194**, 362.

XVII. *Electron Energies resulting from an Electric Field in a Highly Ionized Gas.*

By R. G. GIOVANELLI, M.Sc.*

[Received September 17, 1946.]

1. ELECTRON ENERGIES IN HIGHLY IONIZED GASES.

(a) *Introduction.*

IN normal gases, the current which flows when an electric field is applied is limited, for small fields, by elastic collisions and, as the field increases, by inelastic collisions with neutral atoms. When the field strength is sufficiently great for ionization to occur on collision, the current suddenly increases to a high value and a discharge is said to take place, the mechanism involved being reasonably well known.

Additional ionization is impossible with a completely ionized gas, and so a discharge of the normal type initiated by a critical field strength cannot occur. Nevertheless it will be shown that as the electric field applied to such a gas is increased the current, which is initially limited by elastic collisions between electrons and positive ions, increases rapidly as the field strength reaches a critical value. This is due to a reduction in the cross-section of positive ions for scattering of electrons with increasing electron velocity. The rapid increase of current produced in this way is similar to a discharge and will be referred to as such. These discharges may well be of importance in stars where the ionization is already high.

In the following treatment, which refers only to highly ionized gases, the energies acquired by electrons in an electric field, the variations of the electrical conductivity with field strength and the critical value of the latter for a discharge will be discussed. The analysis is essentially a simplified one, particularly in that no allowance has been made for the distribution of energies and free paths about their means.

(b) *The Electron Drift Velocity.*

Consider a neutral gas consisting of completely ionized atoms all of the same type, the charge on each being Ze . The ionized atoms and free electrons will be assumed initially to be in thermal equilibrium, so that the average energies of electrons and atoms are equal.

If an electric field, E , e.m.u. of infinite extent, be now applied at right angles to a magnetic field H , the electrons will drift through the atoms with a drift velocity which has a component v in the direction of the electric field. Since the atoms are much heavier and therefore slower than the electrons, the main current will be due to the latter. The former will therefore be neglected in so far as current is concerned.

* Communicated by Division of Physics, National Standards Laboratory.

It is well known that the drift velocity of electrons through a gas in the direction of the electric field is given by

$$v = \frac{Ee\tau}{m\left(1 + \frac{e^2 H^2 \tau^2}{m^2}\right)},$$

where τ is the mean time between collisions and m is the electron mass. τ is itself given by λ/u , where λ is the mean free path and u the mean electron velocity between collisions. λ is given by

$$\lambda = \frac{1}{\pi N \Sigma \sigma^2},$$

where σ is the effective collision radius for any type of collision which may occur between an electron and the atoms present, and N is the number of atoms per unit volume.

The collisions of electrons with other electrons may immediately be seen to be of no consequence since if these collisions be elastic, as they presumably are, both the mean energy and momentum will be preserved.

Cowling has discussed collisions between electrons and positive ions, and has shown (1945) that for conditions such as apply in the sun's atmosphere, σ takes the form

$$\sigma = \gamma Z(5750/T) \text{ cm.},$$

T being the electron temperature and Z the mean degree of ionization. He has pointed out that γ is an insensitive function of the electron concentration and temperature. Chapman and Cowling's discussion (1939) yields the result

$$\gamma = \frac{Ze^2}{2 \times 5750k} \left\{ \log_e \left[1 + \left(\frac{4kT}{Ze^2 N_0^{1/3}} \right)^2 \right] \right\}^{\frac{1}{2}},$$

where N_0 is approximately equal to N and k is Boltzmann's constant, although it is not explicitly given in this form. The values of γ over a wide range of conditions are given in Table I., and it is clear that for

TABLE I.

Values of $10^7 \gamma$ for various values of N and T .

$\begin{array}{c} N \\ \text{(per c.c.)} \\ T \end{array}$	10^7	10^9	10^{11}	10^{13}
5,750° K.	6.84	6.35	5.81	5.23
50,000	7.48	7.03	6.55	6.04
500,000	8.10	7.69	7.26	6.80
3,000,000	8.56	8.17	7.76	7.33

conditions in the solar chromosphere and corona, where electron concentrations range from about 10^7 to 10^{12} per c.c., it is safe to regard γ as a constant, equal to 6.0×10^{-7} . Using this value, σ may be written in the form

$$\sigma = \alpha Z / u_1^2 \text{ cm.},$$

where $\alpha = 1.57 \times \text{cm}^3 \cdot 10^9 \text{ sec}^{-2}$ and u_1 is the actual electron velocity. u_1 is closely equal to u , the mean velocity between collisions, unless the increase in energy between collisions be large. This latter occurs only when the electrons are gaining energy rapidly, and the subsequent analysis, for this and other reasons, does not apply in such a case. Replacing u_1 by u , we obtain

$$\lambda = u^4 / \pi \alpha^2 N Z^2 \text{ cm.},$$

and

$$\tau = u^3 / \pi \alpha^2 N Z^2 \text{ sec.}$$

Making the simplifying assumption that, at any given time after the application of the electric field the electron velocities are all the same, it follows that

$$v = \frac{E e u^3}{\pi \alpha^2 N m Z^2 \left(1 + \frac{e^2 H^2 u^6}{\pi^2 \alpha^4 N^2 m^2 Z^4} \right)} \quad \dots \quad (1)$$

(c) *Average Energy lost by an Electron at an Elastic Collision.*

If an electron of mass m , velocity u makes an elastic collision with an atom of mass M and velocity U , the electron and atom being regarded as smooth rigid elastic spheres, then the energy lost by the electron is

$$\frac{2mMs^2}{(m+M)^2} \left(m + \frac{MS}{s} \right) \left(1 - \frac{S}{s} \right)$$

if the components s and S , of the velocities of the electron and atom respectively, along the line of centres immediately before impact are in the same direction, and

$$\frac{2mMs^2}{(m+M)^2} \left(m - \frac{MS}{s} \right) \left(1 + \frac{S}{s} \right)$$

if the two component velocities are in the opposite direction. The mean energy lost by the electron for both types of collision, assumed equally frequent, is thus very closely

$$\frac{2m^2 s^2}{M} \left(1 - \frac{MS^2}{ms^2} \right).$$

The fraction of electrons whose component velocities along the line centres at impact lie between s and $s+ds$ is

$$\frac{dn}{n} = \frac{2sds}{u^2},$$

while for atoms the corresponding fraction is

$$\frac{dN}{N} = \frac{2SdS}{U^2}.$$

Thus the mean energy lost by the electron for all types of collision is

$$\begin{aligned} & \int \frac{2m^2s^2}{M} \left(1 - \frac{MS^2}{ms^2} \right) \frac{dn}{n} \cdot \frac{dN}{N} \\ &= \frac{m^2u^2}{M} \left(1 - \frac{1}{k} \right), \end{aligned}$$

where k is mu^2/MU^2 , i. e. the ratio of the kinetic energy of an electron to that of an atom.

(d) *The Rate of Increase of Electron Energy in Electric Fields.*

After a time τ equal to the mean time between collisions, the gain of energy to the average electron will equal $Eev\tau$, the energy gained from the field between collisions, less $(m^2u^2/M)(1-1/k)$, the energy lost per elastic collision. Thus

$$\frac{dW}{dt} = Eev - \frac{m^2u^2}{M\tau} \left(1 - \frac{1}{k} \right),$$

where we have again replaced u_1 by u , the mean velocity between collisions.

Substituting for v and τ ,

$$\begin{aligned} \frac{dW}{dt} = & \frac{E^2e^2u^3}{\pi\alpha^2mNZ^2 \left(1 + \frac{e^2H^2u^6}{\pi^2\alpha^4N^2m^2Z^4} \right)} \dots \dots \dots (2) \\ & - \frac{\pi\alpha^2m^2NZ^2}{Mu} \left(1 - \frac{u_0^2}{u^2} \right), \end{aligned}$$

where u_0 is the velocity of electrons whose kinetic energy equals the mean kinetic energy of the ions.

(e) *Equilibrium and Discharge Conditions.*

The electron energy reaches a constant value when dW/dt is zero. The equilibrium value of the electron velocity is thus found by equating the right-hand side of the above equation to zero and solving for u .

If E is zero, the electron energy is constant when $u = u_0$, as is to be expected, and this is a stable state; for if u exceeds u_0 the rate of increase of energy is negative and if u is less than u_0 the rate of increase of energy is positive.

We shall now consider cases in which H is zero and E is not zero. The rate at which electrons gain energy is seen to be a function of u_0 , and thus of the kinetic energy of the positive ions. At any time, the average momenta carried by charges of opposite sign will be equal and opposite: thus the drift velocities of positive and negative charges will be in the ratio Zm/M , and electrons gain energy from the field much more rapidly than do positive ions, in the ratio M/Zm . Thus, if an electric field be suddenly applied we may regard u_0 as temporarily constant, and use (2) to ascertain the initial rate of gain of energy to electrons. However, if there be no way

of losing the energy acquired, then the kinetic energy of positive ions gradually increases, u_0 increases and, for any given value of u , dW/dt gradually increases. No stable condition can result, for in such a case u and u_0 would necessarily be equal: but then dW/dt would be positive, and electrons would be gaining energy, contradicting the assumption of stability. Electron and positive ion energies thus increase without limit.

In any real gas of finite extent there will be some means whereby energy can be lost, and the energies of positive ions corresponding to any electric field will have some equilibrium value. We may take as the lower limit for u_0 the value corresponding to the initial temperature of the gas. The appropriate values of the equilibrium electron energies follow immediately, and these are the smallest that can occur in a real gas. We shall now discuss this important case in detail.

If $dW/dt=0$, equation (2) above has two roots for u between u_0 and infinity. The smaller of these roots is stable and the larger unstable, so that if by any means u becomes great enough, dW/dt becomes positive and the electron energy increases without limit. This is, of course, obvious from equation (2) by inspection.

If E is great enough, the above roots become imaginary, and no stable current can exist. The maximum value of E for which a stable current is possible can be shown to be given by

$$E_c = \frac{2\sqrt{3}}{9} \cdot \frac{\pi \alpha^2 m^{3/2} N Z^2}{u_0^2 c M^{1/2}}, \quad (3)$$

which becomes, for hydrogen ions,

$$E_c = 3.95 \times 10^9 \text{ N}/u_0^2 \text{ c.m.u.},$$

or

$$E_c = 1.51 \times 10^{-6} \text{ N c.m.u./cm.}$$

if u_0^2 corresponds to 5750°K . N , the concentration of hydrogen ions, equals n , the electron concentration, if the gas be electrically neutral. As E_c is inversely proportional to u_0^2 , the value of the critical field decreases with increasing temperature. Thus if the rate of energy loss is insufficient to keep the temperature approximately constant, a discharge can occur for a lower value of the electric field than would otherwise be the case.

While the electron energy remains in a state of equilibrium it does not depart far from its thermal energy, *i. e.*, the energy in the absence of the field (see Table II.). In practice random fluctuations will cause electrons to pass from the stable solution of the above equation to the unstable, or discharge, region for values of E somewhat less than given above.

The above discussion also gives conditions which apply when the electric field is parallel to a magnetic field, for in this case it is well known that the motion in the direction of the electric field is not influenced by the magnetic field.

Returning to equation (2), if the magnetic field is perpendicular to the electric field and is not zero, then, for a given electric field, the rate of gain of energy will always be less than for a zero magnetic field and electrons

cannot acquire an unlimited energy. A simple solution for the equilibrium energy is found if

$$\frac{e^2 H^2 u^6}{\pi^2 x^4 N^2 m^2 Z^4}$$

is considerably greater than unity, *e. g.*, if *H* is large, for then (2) becomes

$$\frac{dW}{dt} = \frac{E^2 \pi x^2 N m Z^2}{H^2 u^3} - \frac{\pi x^2 m^2 N Z^2}{M u} \left(1 - \frac{u_0^2}{u^2}\right),$$

TABLE II.

E/N	u/u_0	$(u/u_0)^2$	Conductivity (e.m.u.)
0	1.00	1.00	4.88×10^{-9}
1.0×10^{-6}	1.04	1.09	5.48×10^{-9}
1.1×10^{-6}	1.05	1.11	5.64×10^{-9}
1.2×10^{-6}	1.07	1.14	5.97×10^{-9}
1.3×10^{-6}	1.09	1.18	6.31×10^{-9}
1.4×10^{-6}	1.12	1.25	6.85×10^{-9}
1.45×10^{-6}	1.14	1.30	7.22×10^{-9}
1.48×10^{-6}	1.16	1.35	7.61×10^{-9}
1.51×10^{-6}	1.22	1.50	8.85×10^{-9}

Relative values of electron energies and electrical conductivities with increasing electric field strengths in hydrogen at 5750° K. ($u_0 = 5.11 \times 10^7$ cm./sec.). For values of *E/N* greater than 1.51×10^{-6} elastic collisions cannot limit the electron energy. The above results apply only when *H* is zero.

and when

$$\frac{dW}{dt} = 0,$$

$$\frac{1}{2} m (u^2 - u_0^2) = \frac{M E^2}{2 H^2},$$

giving the increase in energy due to the application of the electric field.

(f) The Electric Current.

The electric current *i* is given by *nev* per cm.², where *n* is the electron concentration, *i. e.* *ZN* if the gas as a whole is neutral. Thus

$$i = \frac{E e^2 u^3}{\pi x^2 m Z \left(1 + \frac{e^2 H^2 u^6}{\pi^2 x^4 N^2 m^2 Z^4}\right)} \dots \dots \dots (4)$$

It is of importance that if *H* be very small the current is approximately proportional to *E* while *E* is below the critical value, as the electron velocity does not depart far from its thermal value. Values of the conductivity in the absence of a magnetic field are given in Table II. Owing to the simplified nature of the approach to this problem, the actual numerical values of the conductivity given here are not as reliable as that given by Cowling's formula for the weak field conductivity, namely 3.7×10^{-9} e.m.u.

They do show, however, the relative variations that take place as the electric field increases.

The above simplified treatment has shown that, provided the field strength is below a certain value, a stable current can flow in a highly ionized gas. Once the field strength exceeds a certain limit, however, the collision cross section, in the absence of a magnetic field, is reduced to such an extent that it can no longer impede the flow of electrons, and in a completely ionized gas in which inelastic collisions are impossible their energies increase to a limit imposed only by the geometrical extent of the field.

(g) *The Effect of Incompletely Ionized Atoms.*

In general the ionization of any gas will be incomplete. However, the cross-section for an elastic collision with a neutral atom is unlikely to be other than a small fraction of that with an ion, and thus if elastic collisions at thermal velocities occur mainly with ions, they may be expected to occur mainly with ions at higher velocities.

If the magnitude of an electric field be such that elastic collisions cannot limit the electron velocity, the only effective slowing-down processes can be those involving excitation or ionization which, however, cannot take place until the electron energy has increased to the excitation energy. The possible limitation of electron velocities by these processes is reasonably well understood, and needs no discussion here.

(h) *The Effect of a Velocity Distribution.*

Distributions of velocities and mean free paths have been omitted throughout, and for simplicity it has been assumed that the initial electron velocities have been identical, although with random orientations.

On introducing a distribution of free paths about a mean, it is clear that the critical nature of E for discharge conditions is somewhat smoothed, for some electrons will acquire high enough velocities to be accelerated further while others will have much smaller velocities. Thus, the current will not be discontinuous when E reaches the critical value, but will increase somewhat more rapidly than was indicated above before E reaches this limit.

2. APPLICATIONS IN THE SUN'S ATMOSPHERE.

(a) *The Conductivity for Weak Fields.*

The sun's atmosphere is highly ionized, and Cowling has pointed out that under normal circumstances elastic collisions of electrons there are mainly with ions. The theory developed above is therefore applicable to this case.

We shall assume that in the chromosphere the gas is solely hydrogen so that $Z=1$, and that $u_0=5.11 \times 10^7$ cm./sec., corresponding to thermodynamic equilibrium velocities at 5750° K. Equilibrium values of u in the absence of a magnetic field, obtained for these conditions by solving (2) for $dW/dt=0$, are given in Table II., these being expressed for convenience in terms of u_0 , the mean electron velocity for thermodynamic equilibrium at 5750° K. The fractional increase in electron energy, $(u/u_0)^2$, is also shown together with the values of the electrical conductivity, *i.e.* the coefficient of E in equation (4). The critical value for E/N is 1.51×10^{-6} ,

above which the current is unstable and the electron velocity cannot be limited by elastic collisions.

(b) *The Electrical Conductivity when Electron Velocities are Limited by Inelastic Collisions.*

When the field strength is sufficiently great and H is zero, elastic collisions with ions will not limit the electron velocity, and the energy will increase either until it be limited by inelastic collisions or, if there are insufficient excitable atoms, by other factors such as the extent of the field. The drift velocity increases with the mean electron velocity, and thus the current also increases. For purposes of illustration, the current will be evaluated for the case where electrons are prevented from acquiring

TABLE III.

	Time taken.	Average drift distance.	Mean drift velocity during time taken to acquire this velocity.
$u=u_0$ Initial condition at instant field is applied.	secs. 0	cm. 0	cm./sec. 4.9×10^5
$n=1.22u_0$ Rate of gain of electron energy is a minimum	1.5×10^{-3}	1.0×10^3	6.7×10^5
$n=1.88 \times 10^8$ cm./sec. Electron energy equals excitation potential of neutral hydrogen.	2.8×10^{-3}	7.6×10^3	2.7×10^6

The above table applies when the gas is hydrogen at 5750° K., the electron concentration is 10^{11} per c.c. and E is 1.6×10^5 e.m.u./cm. For the critical field of 1.51×10^5 e.m.u./cm. the critical drift velocity is 8.5×10^5 cm./sec.

energies greater than 10.1 electron volts ($u=1.88 \times 10^8$ cm./sec.) by exciting the first resonance level in neutral hydrogen atoms, as may happen in the lower chromosphere. As before, the gas will be assumed to be solely hydrogen, the ambient temperature 5750° K.

For given values of E and N equation (1) may be integrated numerically, giving values of u at different times. From this the drift velocity at any time may be calculated. The average distance travelled may be obtained numerically from $\int v dt$, and the mean velocity can thus be found. Some results for the case $N=10^{11}$ per c.c. (corresponding to a level not quite 2000 Kms. above the base of the chromosphere) and $E=1.6 \times 10^5$ e.m.u./cm. are given in Table III. It will be noted that if the field be

gradually increased the current, which is proportional to the drift velocity, increases by a factor of about three times once the critical field is passed, and the conductivity is about five times the conductivity for a very weak field.

If there is an excess of energy remaining after an inelastic collision, the drift velocity and the current will be greater than indicated above.

For a zero magnetic field, the rate of gain of energy may be written as

$$\frac{dW}{dt} = Nf\left(\frac{E}{N}\right),$$

so that when the field is at its critical value, E/N is constant and the rate of gain of energy is proportional to the ion concentration. Thus, for a given fraction of the critical field, the time to acquire a given energy is inversely proportional to the positive ion or electron concentration.

(c) *The Electrical Conductivity in the Upper Chromosphere and Corona.*

In the upper chromosphere and corona, where the ionization is very high, the currents will not be limited by excitation, and if a field greater than the critical field persists, the electron energy will continue to increase until limited by some additional factor. The drift velocity and hence the conductivity thus becomes extremely large.

The author would like to express his thanks to Mr. J. K. Mackenzie for many discussions on various aspects of the above paper, in particular the critical values for E ; and also to the Council for Scientific and Industrial Research for permission to publish this research.

REFERENCES.

COWLING, T. G., 1945, *Proc. Roy. Soc.*, **183**, p. 453.

CHAPMAN, S., and COWLING, T. G., 1939, *The Mathematical Theory of Non-uniform Gases* (Cambridge: University Press).

XVIII. *The Flow behind a Stationary Shock.*

By M. J. LIGHTHILL, Department of Mathematics,
University of Manchester*.

[Received October 26, 1948.]

A LINEARIZED equation of the steady inviscid adiabatic flow of a gas, not assumed perfect, isentropic or isenergetic, is given in a simple form. It is used to study the bending of the bow shock on a supersonic aerofoil due to the incidence of waves from the surface, without assuming that the shock strength is small: the waves are shown to be reflected back along the characteristics and figures for the reflection coefficient are obtained.

* Communicated by Professor S. Goldstein, F.R.S.

For a thin supersonic aerofoil, the method is used to discuss the entropy boundary layer due to inexact leading edge sharpness. Similar ideas will be used in later publications to discuss problems of shock diffraction.

§ 1. THE LINEARIZED EQUATION.

For steady inviscid gas flow the velocity \mathbf{q} , the density ρ and the pressure p satisfy

$$\operatorname{div}(\rho\mathbf{q})=0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\rho\mathbf{q} \cdot \nabla\mathbf{q} + \nabla p = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The pressure is a function of ρ and of s , the entropy: in adiabatic flow s will be constant on each streamline.

If $\mathbf{q} = (u_1 + u, v, w)$, and $u, v, w, \rho - \rho_1$ and $p - p_1$ are small, equations (1) and (2) become approximately

$$u_1 \frac{\partial \rho}{\partial x} + \rho_1 \left(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\rho_1 u_1 \frac{\partial u}{\partial x} + \frac{\partial p}{\partial x} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$\rho_1 u_1 \frac{\partial v}{\partial x} + \frac{\partial p}{\partial y} = 0, \quad \rho_1 u_1 \frac{\partial w}{\partial x} + \frac{\partial p}{\partial z} = 0; \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

and s may be taken independent of x so that

$$\frac{\partial \rho}{\partial x} = \left(\frac{\partial \rho_1}{\partial p_1} \right)_s \frac{dp}{dx} = \frac{1}{a_1^2} \frac{\partial p}{\partial x}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

If now we put

$$u' = \frac{p_1 - p}{\rho_1 u_1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

equations (5) show that

$$\frac{\partial v}{\partial x} - \frac{\partial u'}{\partial y} = 0, \quad \frac{\partial u'}{\partial z} - \frac{\partial w}{\partial x} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

and hence also

$$\frac{\partial}{\partial x} \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \right) = 0; \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

so that (u', v, w) is an irrotational vector field provided that the component of vorticity along the streamlines vanishes on some surface whence all the streamlines spring, and that in the intervening region the assumption of approximate uniformity of flow holds. In Appendix 1 it is shown that in the important case when the surface is a shock, ahead of which the flow is uniform, this result is correct provided that the velocity tangential to the shock is along a principal direction of curvature of the shock. Since this condition covers the two-dimensional case with which §§ 2 and 3 alone deal, it is assumed below that (u', v, w) is irrotational. (Even if it is not, u' still satisfies equation (10) below, and v, w can be obtained therefrom by (8).)

Putting $(u', v, w) = \nabla\phi$, equations (3), (4), (6) and (7) combine to give

$$\nabla^2\phi = \frac{u_1^2}{a_1^2} \frac{\partial^2\phi}{\partial x^2} = M_1^2 \frac{\partial^2\phi}{\partial x^2}, \quad \dots \quad (10)$$

showing that the linearized equation of isentropic flow still holds for anisentropic anisenergetic flow if u be replaced by u' .

On any streamline the sum of the kinetic energy $\frac{1}{2}q^2$ and the enthalpy ("total heat") i , per unit mass, will take (in steady flow) a constant value, say

$$\frac{1}{2}(u_1+u)^2 + \frac{1}{2}v^2 + \frac{1}{2}w^2 + i = \frac{1}{2}u_1^2 + i_1 + h, \quad \dots \quad (11)$$

so that h measures variations in energy on different streamlines. Now if $s=0$ and the temperature is T_1 in the undisturbed state,

$$i - i_1 = T_1 s + (p - p_1)/\rho_1.$$

Hence (11) is approximately

$$u' = u + (T_1 s - h)/u_1, \quad \dots \quad (12)$$

which shows how u' reduces to u when $s=h=0$.

§ 2. FLOW BEHIND A SHOCK.

If the above theory is to be applied to the two-dimensional flow produced by the passage of a uniform stream through a shock, the shock must be of nearly uniform strength, since the theory requires nearly uniform flow behind it: and the strength should not be very small as entropy variations would then be insignificant compared with velocity variations. (But the flow is nearly uniform in *some* neighbourhood of any point on a shock, so our results will be locally correct even if the shock as a whole is not of nearly uniform strength.)

Let velocity density and pressure ahead of the shock be q_0 , ρ_0 , p_0 and let the shock be at an angle $\alpha_0 + \alpha$ to this uniform stream: suppose that when $\alpha=0$ the flow behind the shock has velocity $(u_1, 0)$, the x -axis being at angle β to the shock, and density and pressure ρ_1 , p_1 ; these quantities being replaced by (u_1+u, v) , ρ , p when $\alpha \neq 0$. Then the Rankine shock equations are

$$\rho_0 q_0 \sin(\alpha_0 + \alpha) = \rho[(u_1+u) \sin(\beta + \alpha) - v \cos(\beta + \alpha)], \quad \dots \quad (13)$$

$$q_0 \cos(\alpha_0 + \alpha) = (u_1+u) \cos(\beta + \alpha) + v \sin(\beta + \alpha), \quad \dots \quad (14)$$

$$p_0 + \rho_0 q_0^2 \sin^2(\alpha_0 + \alpha) = p + \rho[(u_1+u) \sin(\beta + \alpha) - v \cos(\beta + \alpha)]^2, \quad (15)$$

$$\frac{1}{2}q_0^2 + i_0 = \frac{1}{2}[(u_1+u)^2 + v^2] + i; \quad \dots \quad (16)$$

and assuming that α , u , v , $p-p_1$ and $\rho-\rho_1$ are small and vanish simultaneously, their mutual factors of proportionality can be deduced. In particular we can write $u' = -U\alpha$, $v = V\alpha$ (where u' is as defined in (7)).

If the shock is produced by an aerofoil in the stream (a wedge, if it is to be of uniform strength throughout), any curvature of the surface will cause disturbances propagated along characteristics $y - x \tan \mu_1 = \text{constant}$,

where μ_1 is the Mach angle $\sin^{-1}(a_1/u_1)$. This disturbance flow will be given by a potential $\phi=f(y-x \tan \mu_1)$ —in the sense of § 1—where $f'(y-x \tan \mu_1)$ on the surface is the value of v , i. e. u_1 multiplied by the slope, and is hence known. The disturbances reflected from the shock (if there are any) will have $\phi=g(y+x \tan \mu_1)$. To determine g we have the approximate conditions at the shock $y=x \tan \beta$ and $\partial\phi/\partial x=-UV^{-1}\partial\phi/\partial y$. Hence

$$\begin{aligned} & -\tan \mu_1 f'(x \tan \beta - x \tan \mu_1) + \tan \mu_1 g'(x \tan \beta + x \tan \mu_1) \\ & = -UV^{-1}(f'(x \tan \beta - x \tan \mu_1) + g'(x \tan \beta + x \tan \mu_1)), \quad (17) \end{aligned}$$

so that $g'(x \tan \beta + x \tan \mu_1) = -Rf'(x \tan \beta - x \tan \mu_1)$, where R , the reflection coefficient, is given by

$$R = \frac{U - V \tan \mu_1}{U + V \tan \mu_1}. \quad (18)$$

If the strength of a disturbance is measured by the upwash (v/u_1) which it produces, we can say that in the wedge flow any disturbance is reflected back along the other characteristics, simply multiplied by the constant factor $-R$. The same conclusion is true *locally* at a point of any shock wave with uniform flow ahead of it: but, with no approximate uniformity behind it, the disturbance in (u', v) will not in general remain constant along characteristics.

In Appendix 2, R is obtained explicitly (for air) in terms of M_0 (the Mach number ahead of the shock) and p_1/p_0 . The Table below shows its behaviour for $M_0=2$, a typical "ordinary supersonic" Mach number, for different values of p_1/p_0 ; the Mach number M_1 behind the shock, and the deflection δ , being also tabulated. Reflection of waves from the shock at this Mach number is seen to be on a small scale. But for "hypersonic" M_0 , it plays an important rôle; and the appendix shows that R is near -1 for the shock waves produced by an aerofoil at high Mach numbers.

p_1/p_0	M_1	δ	R
1.5	1.732	7.49°	0.00194
2	1.521	13.14°	0.00789
2.5	1.333	17.58	0.02035
3	1.155	20.92	0.03779

The shape of the shock can also be found, since

$$\alpha = \frac{1}{V} \frac{\partial\phi}{\partial y} = \frac{1-R}{V} f'(y-x \tan \mu_1) = \frac{2 \tan \mu_1}{U+V \tan \mu_1} f'(y-x \tan \mu_1), \quad (19)$$

from which it follows that the shape of the shock will reproduce the shape of the wedge (to the first order), the slope being altered by the constant factor

$2u_1 \tan \mu_1 / (V \tan \mu_1 + U)$ —this factor is of order 1, its value for weak shocks—and the scale increased by the factor $\sin \mu_1 / \sin (\mu_1 - \beta)$. This fact should be well within the limits of experimental observation, though its precision will be blurred owing to the non-parallelism of the true characteristics.

§ 3. EFFECT OF INEXACT LEADING EDGE SHARPNESS ON THE FLOW PAST A THIN SUPERSONIC AEROFOIL

If (as must happen) a thin supersonic aerofoil has a leading edge not exactly sharp, the shock will be much stronger near this point than elsewhere (perhaps detached): and the resulting comparatively large entropy variations will be propagated along the streamlines to form an "entropy boundary layer". Outside a neighbourhood of the leading edge the flow will be approximately uniform and of "progressive wave" character, but with entropy variations. The ordinary theory of thin supersonic aerofoils can then be applied using u', v instead of u, v and we may deduce that u' is unaffected by the entropy boundary layer and that on the surface $u' = -v \tan \mu_1$ as usual. Thus the pressure on the surface is unaffected (as in an ordinary boundary layer), while the velocity u on the surface is considerably altered as (12) shows. Effectively then the bluntness of the leading edge, as well as producing local effects, increases the vorticity near the whole surface, above that due to viscous stress and heat conduction, but does not affect the pressure.

APPENDIX 1.

Behind a shock the component of vorticity along a streamline, multiplied by the velocity, is

$$\mathbf{q} \cdot \text{curl } \mathbf{q} = u \left(\frac{\partial w}{\partial y} - \frac{\partial v}{\partial z} \right) + v \left(\frac{\partial u}{\partial z} - \frac{\partial w}{\partial x} \right) + w \left(\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right). \quad (19)$$

Choose an origin on the shock, with x -axis perpendicular to it and y, z -axes along principal directions of curvature at the origin (with associated curvatures κ_1, κ_2). At the origin only u differs from its value in the uniform flow ahead of the shock: at $(dy, 0)$ v but not w differs; at $(0, dz)$ w but not v ; hence $\partial w / \partial y = \partial v / \partial z = 0$. Further

$$u \frac{\partial v}{\partial x} + v \frac{\partial v}{\partial y} + w \frac{\partial v}{\partial z} + \frac{1}{\rho} \frac{\partial p}{\partial y} = u \frac{\partial w}{\partial x} + v \frac{\partial w}{\partial y} + w \frac{\partial w}{\partial z} + \frac{1}{\rho} \frac{\partial p}{\partial z} = 0, \quad (20)$$

so that

$$\begin{aligned} \mathbf{q} \cdot \text{curl } \mathbf{q} &= v \left(\frac{\partial u}{\partial z} + \frac{w}{u} \frac{\partial w}{\partial z} + \frac{1}{\rho u} \frac{\partial p}{\partial z} \right) - w \left(\frac{\partial u}{\partial y} + \frac{v}{u} \frac{\partial v}{\partial y} + \frac{1}{\rho u} \frac{\partial p}{\partial y} \right) \\ &= \frac{T}{u} \left(-v \frac{\partial s}{\partial z} + w \frac{\partial s}{\partial y} \right). \end{aligned} \quad (21)$$

But the entropy behind the shock, with given pressure and density ahead of it, depends only on the normal component of velocity q_n . At $(dy, 0)$ this is $u_0 - v\kappa_1 dy$, so that $\partial s/\partial y = -(ds/dq_n)v\kappa_1$. Similarly

$$\partial s/\partial z = -(ds/dq_n)w\kappa_2$$

and hence

$$\mathbf{q} \cdot \text{curl } \mathbf{q} = \frac{Tw}{u} \frac{ds}{dq_n} (\kappa_2 - \kappa_1), \quad . \quad . \quad . \quad . \quad . \quad (22)$$

showing that the component of vorticity in the direction of flow vanishes only if the velocity tangential to the shock is in a principal direction of curvature (with the usual convention that if $\kappa_1 = \kappa_2$ all directions are principal directions).

APPENDIX 2.

From equation (15)

$$2\rho_0 q_0^2 \sin \alpha_0 \cos \alpha_0 \cdot \alpha = p - p_1 + 2\rho_1 u_1 \sin \beta (u_1 \cos \beta \cdot \alpha + u \sin \beta - v \cos \beta) \\ + (\rho - \rho_1) u_1^2 \sin^2 \beta, \quad . \quad . \quad . \quad . \quad . \quad (23)$$

whence, by (13) and (14) with $\alpha = 0$,

$$p - p_1 + 2\rho_1 u_1 \sin \beta (u \sin \beta - v \cos \beta) + (\rho - \rho_1) u_1^2 \sin^2 \beta = 0. \quad . \quad (24)$$

Also by (16), taking $i = 7p/2\rho$, its value for air,

$$0 = uu_1 + \frac{7}{2} \left(\frac{p - p_1}{\rho_1} - \frac{p_1(\rho - \rho_1)}{\rho_1^2} \right). \quad . \quad . \quad . \quad . \quad . \quad (25)$$

But it is well known that equations (13) to (16) imply that

$$p(6\rho_0 - \rho) = p_0(6\rho - \rho_0), \quad . \quad . \quad . \quad . \quad . \quad (26)$$

whence

$$(p - p_1)(6\rho_0 - \rho_1) - p_1(\rho - \rho_1) = 6p_0(\rho - \rho_1), \quad . \quad . \quad (26)$$

Hence (24) becomes

$$2\rho_1 u_1 \sin \beta \cos \beta \cdot v \\ = p - p_1 + (\rho - \rho_1) u_1^2 \sin^2 \beta + 2\rho_1 \sin^2 \beta \cdot \frac{7}{2} \left(\frac{p_1(\rho - \rho_1)}{\rho_1^2} - \frac{p - p_1}{\rho_1} \right) \\ = (p - p_1) \left[1 - 7 \sin^2 \beta + \left(u_1^2 + \frac{7p_1}{\rho_1} \right) \sin^2 \beta \frac{6\rho_0 - \rho_1}{6p_0 + p_1} \right]. \quad . \quad . \quad . \quad . \quad (28)$$

With $p_1/p_0 = y$, $\rho_1/\rho_0 = x = (6y + 1)/(6 + y)$, and with U, V as defined in § 2, we deduce (using $u_1^2 + 7p_1/\rho_1 = q_0^2 + 7p_0/q_0$) that

$$\frac{V}{U} = \frac{1}{2} \tan \beta \left[\text{cosec}^2 \beta - 7 + \frac{7}{5} (M_0^2 + 5) \frac{6 - x}{6 + y} \right]. \quad . \quad . \quad . \quad . \quad (29)$$

But from (13) and (14) with $\alpha=0$ we have $\rho_0^2 q_0^2 - u_1^2(\rho_1^2 \sin^2 \beta + \rho_0^2 - \rho_0^2 \sin^2 \beta)$, so that

$$\begin{aligned} \operatorname{cosec}^2 \beta &= \frac{u_1^2(\rho_1^2 - \rho_0^2)}{\rho_0^2(q_0^2 - u_1^2)} = \frac{u_1^2(\rho_1^2 - \rho_0^2)}{7\rho_0^2(p_1/\rho_1 - p_0/\rho_0)} = \frac{M_1^2(x^2 - 1)}{5(1 - x/y)} \\ &= \frac{M_1^2 y}{5(6+y)} \frac{(6y+1)^2 - (6+y)^2}{y(6+y) - (6y+1)} = \frac{7y}{6+y} M_1^2; \quad \dots \quad (30) \end{aligned}$$

and hence

$$\frac{V}{U} \tan \mu_1 = \frac{1}{2} \left(\frac{7y}{6+y} M_1^2 - 1 \right)^{-\frac{1}{2}} (M_1^2 - 1)^{-\frac{1}{2}} \left[\frac{7y}{6+y} M_1^2 - 7 + \left(\frac{7}{6+y} \right)^2 (M_0^2 + 5) \right], \quad \dots \quad (31)$$

from which R is found as a function of M_0 and y above by (18) and the known result

$$M_1^2 = \frac{M_0^2(6y+1) - 5(y^2 - 1)}{y(6+y)} \quad \dots \quad (32)$$

For hypersonic M_0 and given deflection $\delta < 45.58^\circ$ produced by the shock we have asymptotically $y \sim Y M_0^2$, with

$$0 < Y < 1, \quad \tan \delta = \frac{5Y}{7-5Y} \sqrt{\frac{7}{6Y}} - 1, \quad M_1^2 \sim \frac{6-5Y}{Y} \quad \dots \quad (33)$$

Hence

$$\begin{aligned} \frac{V}{U} \tan \mu_1 &\sim \frac{1}{2} \left(7 \frac{6-5Y}{Y} - 1 \right)^{-\frac{1}{2}} \left(\frac{6-5Y}{Y} - 1 \right)^{-\frac{1}{2}} \left[7 \frac{6-5Y}{Y} - 7 + \frac{49}{Y^2} \right] \\ &= \frac{7(7+6Y-6Y^2)}{12Y(7-6Y)^{\frac{1}{2}}(1-Y)^{\frac{1}{2}}}, \quad \dots \quad (34) \end{aligned}$$

a rather large quantity when $0 < Y < 1$. Thus R is near -1 for hypersonic flow.

XIX. Note on a new Form of the Solution of Reynolds' Equation for Michell Rectangular and Sector-shaped Pads.

By Mrs. W. L. Wood, M.A. *

[Received November 29, 1947.]

1. MICHELL'S method of solving Reynolds' equation for finite bearing pads is given in *Zeitschrift für Mathematik und Physik*, 1905, and recounted by H. M. Martin in *Engineering*, February, 1920. The solution is given in the form of a series which is often unwieldy. This note shows that it can be expressed more conveniently in a combination of Bessel functions with the Struve function for imaginary argument, tables of which were published in the October, 1946 issue of the *Journal of Mathematics and Physics*.

* Communicated by the Author,

The formulæ for load carried, centre of pressure and flow of lubricant are derived, and their corresponding asymptotic forms are given, as these are often sufficient for cases of practical interest. No account is taken here of the variation of viscosity with temperature.

2. The equation for the rectangular pad is

$$\frac{\partial}{\partial x} \left(h^3 \frac{\partial p}{\partial x} \right) + \frac{\partial}{\partial y} \left(h^3 \frac{\partial p}{\partial y} \right) = 6\eta U \frac{\partial h}{\partial x},$$

with the usual notation,

$h = h_0 \left(1 - \frac{bx}{l} \right)$ and the boundary conditions are $p=0$ at $x=0, l$ and $y=0, \lambda l$.

The equation is non-dimensionalised by writing $x = lx'$, $y = \frac{\lambda l}{\pi} y'$,

$p = \frac{6\eta U b l}{h_0^2} p'$. It then becomes

$$\frac{\partial}{\partial x'} \left[(1 - bx')^3 \frac{\partial p'}{\partial x'} \right] + \frac{\pi^2}{\lambda^2} (1 - bx')^3 \frac{\partial^2 p'}{\partial y'^2} = -1 = -\frac{4}{\pi} \sum_{n=0}^{\infty} \frac{\sin(2n+1)y'}{2n+1}.$$

The solution is

$$p' = \sum_{n=0}^{\infty} \frac{4}{\lambda b^3} \frac{w_n(z)}{z} \sin(2n+1)y',$$

where

$$z = \frac{\pi(2n+1)}{\lambda b} (1 - bx'),$$

and

$$z^2 \frac{d^2 w_n}{dz^2} + z \frac{dw_n}{dz} - (1 + z^2) w_n = -1. \quad \dots \quad (1)$$

Moreover, if

$$z_1 = \frac{\pi(2n+1)(1-b)}{\lambda b}, \quad z_2 = \frac{\pi(2n+1)}{\lambda b},$$

$$w_n(z_1) = w_n(z_2) = 0.$$

3. The equation for the sector pad is

$$\frac{\partial}{\partial \theta} \left(h^3 \frac{\partial p}{\partial \theta} \right) + r \frac{\partial}{\partial r} \left(h^3 r \frac{\partial p}{\partial r} \right) = 6\eta \omega r^2 \frac{\partial h}{\partial \theta},$$

and putting $\theta = \theta_0 \theta'$,

$$h = h_0(1 - b\theta') \quad (\text{the usual approximation}),$$

$$r = r_0 e^{a\theta'},$$

$$p = \frac{6\omega r_0^2 \theta_0 b \eta}{h_0^2} \cdot p',$$

$$z = \frac{n\theta_0}{ab} (1 - b\theta'),$$

the solution is

$$p' = \Sigma H_n \cdot \frac{n\theta_0}{ab^3} \cdot \frac{w_n(z)}{z} \sin nr',$$

where

$$H_n = \frac{2n}{\pi} \left[\frac{1 + (-)^{n+1} e^{2a\pi}}{n^2 + 4a^2} \right],$$

and $w_n(z)$ is again given by equation (1).

Since $p' = 0$ when $r' = 0, \pi$, hence $a = \frac{1}{\pi} \log \frac{r_1}{r_0}$; and $w_n(z_1) = w_n(z_2) = 0$ where in this case

$$z_1 = \frac{n\theta_0}{ab} (1-b), \quad z_2 = \frac{n\theta_0}{ab}.$$

The solution of equation (1) is

$$w_n(z) = A_n I_1(z) + B_n K_1(z) + \frac{\pi}{2} L_{-1}(z).$$

where $I_1(z)$ and $K_1(z)$ are Bessel functions with the usual notation, and $L_{-1}(z)$ is the Struve function for imaginary argument, order -1 .

A_n, B_n are given by $w_n(z_1) = w_n(z_2) = 0$.

4. The load carried by the rectangular pad

$$W_R = \int_0^l \int_0^{\frac{\lambda}{\pi}} p \, dx \, dy.$$

Writing

$$\frac{h_0^2 \pi}{6\eta U b l^3 \lambda} \cdot W_R = W'_R$$

the non-dimensional load, it can easily be seen that

$$\begin{aligned} W'_R &= \Sigma \frac{8}{b^3 \pi (2n+1)^2} \int_{z_1}^{z_2} \frac{w_n(z)}{z} \, dz \\ &= \Sigma \frac{8}{b^3 \pi (2n+1)^2} C_{nR}, \text{ say.} \end{aligned}$$

Similarly the load carried by the sector pad

$$W_S = \int_0^{\theta_0} \int_{r_1}^{r_0} r p \, dr \, d\theta,$$

and

$$\begin{aligned} W'_S &= W_S \cdot \frac{h_0^2}{6\omega r_0^4 \theta_0^2 b \eta} \\ &= \Sigma \frac{\pi a H_n^2}{2b^3} \int_{z_1}^{z_2} \frac{w_n(z)}{z} \, dz \\ &= \Sigma \frac{\pi a H_n^2}{2b^3} C_{nS} \text{ say.} \end{aligned}$$

Let

$$C_n = \int^z \frac{1}{z} \left[A_n I_1(z) + B_n K_1(z) + \frac{\pi}{2} L_{-1}(z) \right] dz.$$

This can be calculated with the help of the formulæ

$$\int^z \frac{I_1(z)}{z} dz = \frac{\pi z}{2} [L_{-1}(z) I_{-2}(z) - L_{-2}(z) I_{-1}(z)],$$

$$\int^z \frac{K_1(z)}{z} dz = -\frac{\pi z}{2} [L_{-1}(z) K_{-2}(z) + L_{-2}(z) K_{-1}(z)].$$

The coefficient of friction for the rectangular pad on the frictional moment for the sector pad can be obtained easily once W'_R or W'_S is known.

For the rectangular pad the centre of pressure (\bar{x}, \bar{y}) is given by $\bar{y} = \lambda l/2$ and $W_R \bar{x} = \int_0^l \int_0^{\lambda} p x dx dy$, i. e.

$$\frac{\bar{x}}{l} W'_R = \frac{1}{b} W'_R - \frac{8\lambda}{\pi^2 b^3} \Sigma \frac{1}{(2n+1)^3} \int_{z_1}^{z_2} w_n(z) dz,$$

where

$$\int_{z_1}^{z_2} w_n(z) dz = \left[A_n I_0(z) - B_n K_0(z) + \frac{\pi}{2} L_0(z) \right]_{z_1}^{z_2}.$$

The quantity of lubricant flowing across a transverse section of the rectangular pad,

$$Q_x = \frac{U h \lambda l}{2} - \int_0^{\lambda} \frac{h^3}{12 \eta} \frac{\partial p}{\partial x} dy.$$

Non-dimensionally

$$Q'_x = \frac{Q_x \cdot \lambda b^2}{4 h_0 U l} = (1 - b x') \cdot \frac{\lambda^2 b^2}{8} + (1 - b x')^3 \Sigma \frac{d}{dz} \left(\frac{w_n(z)}{z} \right).$$

Similarly for the sector pad

$$Q'_\theta = \frac{Q_\theta \cdot a b^2}{h_0 w r_0^2 \theta_0^2} = (1 - b \theta') \left[\left(\frac{r_1}{r_0} \right)^2 - 1 \right] \frac{a b^2}{4 \theta_0^2} + (1 - b \theta')^3 \Sigma H_n \frac{d}{dz} \left(\frac{w_n(z)}{z} \right)$$

and

$$\frac{d}{dz} \left(\frac{w_n(z)}{z} \right) = \frac{1}{z} \left[A_n I_2(z) - B_n K_2(z) + \frac{\pi}{2} L_{-2}(z) \right].$$

The entry flow is $(Q'_x)_{z=z_2}$ or $(Q'_\theta)_{z=z_2}$. Exit flow is $(Q'_x)_{z=z_1}$ or $(Q'_\theta)_{z=z_1}$.

5. In many practical cases the values of z_1 and z_2 are sufficiently large for the asymptotic forms of the above expressions to be used.

Now

$$L_\nu(z) \sim I_{-\nu}(z) + \frac{1}{\pi} \Sigma_{m=0} \frac{(-)^{m+1} \Gamma(m + \frac{1}{2})}{\Gamma(\nu + \frac{1}{2} - m)} \left(\frac{2}{z} \right)^{1+2m-\nu}$$

so that

$$L_0(z) \sim I_0(z) - \frac{2}{\pi z} s_0(z),$$

$$L_{-1}(z) \sim I_1(z) + \frac{2}{\pi z^2} s_1(z),$$

$$L_{-2}(z) \sim I_2(z) - \frac{6}{\pi z^3} s_2(z),$$

where $s_\nu(z)$ is $O(1)$. Moreover,

$$I_\nu(z) \sim \frac{e^z}{\sqrt{2\pi z}} p_\nu,$$

where

$$p_\nu = 1 - \frac{(4\nu^2 - 1^2)}{1! 8z} + \frac{(4\nu^2 - 1^2)(4\nu^2 - 3^2)}{2! (8z)^2} - \dots$$

and

$$K_\nu(z) \sim e^{-z} \sqrt{\frac{\pi}{2z}} q_\nu,$$

where

$$q_\nu = 1 + \frac{(4\nu^2 - 1^2)}{1! 8z} + \frac{(4\nu^2 - 1^2)(4\nu^2 - 3^2)}{2! (8z)^2} + \dots$$

Writing I_{11} for $I_1(z_1)$ etc., A_n , B_n are given by

$$\frac{2}{\pi} A_n = \frac{K_{11} L_{-12} - K_{12} L_{-11}}{I_{11} K_{12} - I_{12} K_{11}},$$

$$\frac{2}{\pi} B_n = \frac{I_{12} L_{-11} - I_{11} L_{-12}}{I_{11} K_{12} - I_{12} K_{11}}.$$

Hence

$$\frac{2}{\pi} A_n \sim -1 - 2 \frac{\sqrt{\frac{2}{\pi}} \frac{e^{-z_2} s_{12}}{z_2^{3/2} p_{12}} \left[\frac{1 - e^{-(z_1 - z_1)} \frac{s_{11} q_{12}}{s_{12} q_{11}} \left(\frac{z_2}{z_1} \right)^{\frac{3}{2}} \right]}{1 - e^{-2(z_1 - z_1)} \frac{p_{11} q_{12}}{p_{12} q_{11}}},$$

i. e.,

$$A_n \sim -\frac{\pi}{2} - 2 \sqrt{\frac{\pi}{2}} \cdot \frac{e^{-z_2} s_{12}}{z_2^{3/2} p_{12}}, \quad \text{since } z_2 > z_1,$$

and

$$B_n \sim -\sqrt{\frac{2}{\pi}} \frac{e^{z_1} s_{11}}{z_1^{3/2} q_{11}} \frac{\left[1 - e^{-(z_1 - z_1)} \frac{p_{11} s_{12}}{p_{12} s_{11}} \left(\frac{z_1}{z_2} \right)^{3/2} \right]}{1 - e^{-2(z_1 - z_1)} \frac{p_{11} q_{12}}{p_{12} q_{11}}},$$

i. e.,

$$B_n \sim -\sqrt{\frac{2}{\pi}} \frac{e^{z_1} s_{11}}{z_1^{3/2} q_{11}}.$$

Hence

$$\frac{1}{z} \left[A_n I_1(z) + \frac{\pi}{2} L_{-1}(z) \right] \sim \frac{1}{z} \left[\frac{1}{z^2} s_1 - 2 \sqrt{\frac{\pi}{2}} \cdot \frac{e^{-z} s_{12}}{z_2^{3/2} p_{12}} I_1(z) \right],$$

and

$$\int_{z_1}^{z_2} \frac{I_1(z)}{z} dz \sim \left[\frac{e^z}{\sqrt{2\pi} \cdot z^{3/2}} (s_1 p_2 + \frac{3}{z} s_2 p_1) \right]_{z_1}^{z_2},$$

also

$$\int_{z_1}^{z_2} \frac{K_1(z)}{z} dz \sim \left[-\frac{\pi z}{2} \left(I_1(z) K_2(z) + I_2(z) K_1(z) + \sqrt{\frac{2}{\pi}} \cdot \frac{e^{-z} s_1 q_2}{z^{5/2}} \right. \right. \\ \left. \left. - 3 \sqrt{\frac{2}{\pi}} \cdot \frac{e^{-z} s_2 q_1}{z^{7/2}} \right) \right]_{z_1}^{z_2}.$$

$$\text{But } z(I_1(z)K_2(z) + I_2(z)K_1(z)) = 1,$$

$$\int_{z_1}^{z_2} \frac{K_1(z)}{z} dz \sim \left[\sqrt{\frac{\pi}{2}} \cdot \frac{e^{-z}}{z^{3/2}} \left(\frac{3}{z} s_2 q_1 - s_1 q_1 \right) \right]_{z_1}^{z_2}.$$

Hence

$$[C_n]_{z_1}^{z_2} \sim \int_{z_1}^{z_2} \frac{s_1}{z^3} dz - \frac{e^{-z} s_{12}}{z_2^{3/2} p_{12}} \left[\frac{e^z}{z^{3/2}} \left(s_1 p_2 + \frac{3}{z} s_2 p_1 \right) \right]_{z_1}^{z_2} \\ - \frac{e^{z_1} s_{11}}{z_1^{3/2} q_{11}} \left[\frac{e^{-z}}{z^{3/2}} \left(\frac{3}{z} s_2 q_1 - s_1 q_2 \right) \right]_{z_1}^{z_2},$$

i. e.,

$$[C_n]_{z_1}^{z_2} \sim \left[-\frac{1}{2z^2} - \frac{3}{4z^4} - \frac{3^2 \cdot 5}{6z^6} \dots \right]_{z_1}^{z_2} \\ - \frac{s_{12}}{z_2^3 p_{12}} \left(s_{12} p_{22} + \frac{3}{z_2} s_{22} p_{12} \right) - \frac{s_{11}}{z_1^3 q_{11}} \left(s_{11} q_{21} - \frac{3}{z} s_{21} q_{11} \right).$$

When z_1, z_2 are sufficiently large this form can be used to evaluate W'_R and W'_S .

The asymptotic form for $\int^z w_n(z) dz$, i.e.

$$A_n I_0(z) - B_n K_0(z) + \frac{\pi}{2} L_0(z),$$

which occurs in the equation for the centre of pressure of the rectangular pad, is

$$-\frac{1}{z} s_0 - \frac{e^{z-z_2}}{z_2^{3/2} z^{1/2}} \frac{s_{12} p_0}{p_{12}} + \frac{e^{z_1-z}}{z_1^{3/2} z^{1/2}} \frac{s_{11}}{q_{11}} q_0,$$

i. e.,

$$\int_{z_1}^{z_2} w_n(z) dz \sim \frac{1}{z_1} s_{01} - \frac{1}{z_2} s_{02} - \frac{s_{12} p_{02}}{z_2^2 p_{12}} - \frac{s_{11} q_{01}}{z_1^2 q_{11}}.$$

Moreover, the expression occurring in Q'_x and Q'_θ ,

$$\frac{d}{dz} \left(\frac{w_n(z)}{z} \right) = \frac{1}{z} \left[A_n I_2(z) - B_n K_2(z) + \frac{\pi}{2} L_{-2}(z) \right] \\ \sim \frac{1}{z} \left[-\frac{3}{z^3} s_2 - \frac{e^{z-z_2} s_{12} p_2}{z_2^{3/2} z^{1/2} p_{12}} + \frac{e^{z_1-z} s_{11} q_2}{z_1^{3/2} z^{1/2} q_{11}} \right],$$

i. e.,

$$\left[\frac{d}{dz} \left(\frac{w_n(z)}{z} \right) \right]_{z=z_1} \sim \left[-\frac{3}{z_1^4} s_{21} + \frac{s_{11} q_{21}}{z_1^3 q_{11}} \right],$$

and

$$\left[\frac{d}{dz} \left(\frac{w_n(z)}{z} \right) \right]_{z=z_2} \sim \left[-\frac{3}{z_2^4} s_{22} - \frac{s_{12} p_{22}}{z_2^3 p_{12}} \right].$$

These forms can be used for evaluating entry and exit flows when z_1, z_2 are sufficiently large.

XX. *A Note on Minimum Integrals in Field Theory.*

By C. W. KILMISTER and B. H. CHIRGWIN *.

[Received February 3, 1948.]

1. INTRODUCTION.

THOMSON'S theorem in electrostatics asserts that the energy

$$U = \frac{1}{8\pi} \iiint \epsilon E^2 d\tau,$$

where ϵ is the dielectric constant, is a minimum for the actual field, the integral being taken over the whole of space.

In current electricity there is a corresponding result that, if \mathbf{J} is the current density vector and \mathbf{E}' represents the applied electromotive intensity, the integral

$$H = \iiint (\mathbf{J}^2/\sigma - 2\mathbf{J} \cdot \mathbf{E}') d\tau,$$

where σ is the conductivity, has a minimum value for the actual current distribution. In the absence of batteries ($\mathbf{E}'=0$) this states that the rate of generation of heat is a minimum.

In an incompressible fluid the energy

$$T = \frac{1}{2} \iiint \rho \mathbf{u}^2 d\tau,$$

where \mathbf{u} is the velocity vector and ρ the mass density, has a minimum value subject to the prescribed boundary conditions, when the motion is irrotational.

* Communicated by the Authors.

The purpose of this note is to point out that all these theorems, which are proved separately (Thomson, Gibbs etc.) as a rule, are special results of a more general field theorem; the theorem is also extended to magnetic fields, verifying a result obtained by Livens (1945, 1947) in another way.

2. STATEMENT OF THE THEOREM.

Consider a region of space, bounded by a number of surfaces, S_1, S_2, \dots, S_p , and which may, if necessary, extend to infinity. Let \mathbf{A} represent a vector field defined in this region, which, together with its derivatives, is finite and continuous. Where the region goes to infinity \mathbf{A} tends to zero sufficiently quickly to make the integrals we shall meet convergent. If $|\mathbf{A}| \rightarrow 0$ as $1/R^2$, this is sufficient.

The theorem concerns conditions for a minimum value of the integral

$$I = \iiint (\mathbf{A} \cdot \boldsymbol{\varphi} \cdot \mathbf{A} + 2\mathbf{C} \cdot \mathbf{A}) d\tau, \quad (1)$$

where $\boldsymbol{\varphi}$ is a prescribed, symmetric, dyadic point function whose components are everywhere finite and at any point form the coefficients of a definite quadratic form (which we shall take as positive definite), and \mathbf{C} is a prescribed finite vector point function. (If $\boldsymbol{\varphi}$ is negative definite our minima are all maxima.)

The theorem falls into two parts according to the way in which \mathbf{A} is defined in terms of its sources or vortices.

(a) (i) Firstly \mathbf{A} is defined from its sources thus

$$\operatorname{div} \mathbf{A} = \mathbf{M} \quad \mathbf{A} \cdot \mathbf{n} = \mathbf{N}, \quad (2)$$

where \mathbf{M} is a prescribed scalar function of position and \mathbf{N} is the prescribed normal component of \mathbf{A} on the boundary (\mathbf{n} is the unit normal drawn outwards from the region of definition of \mathbf{A}). These, for consistency, must be subject to the condition that

$$\iiint \mathbf{M} d\tau = \sum_{i=1}^p \iint \mathbf{N}_i dS_i.$$

When this is so the necessary and sufficient condition for I to have a minimum is

$$\operatorname{curl} (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) = 0. \quad (3)$$

(ii) If the sources of \mathbf{A} are less rigidly defined thus

$$\operatorname{div} \mathbf{A} = \mathbf{M} \quad \iint \mathbf{A} \cdot \mathbf{n} dS_i = \mathbf{E}_i \quad (i=1, 2, \dots, p), \quad (4)$$

i. e. if only the integrated normal component of \mathbf{A} over each bounding surface is specified in addition to \mathbf{M} , there is an additional condition on \mathbf{A} . The necessary and sufficient conditions for I to have a minimum are now

$$\operatorname{curl} (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) = 0 \quad (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) \times \mathbf{n} = 0. \quad (5)$$

Here of course we must have $\iiint \mathbf{M} d\tau = \sum_{i=1}^p \mathbf{E}_i.$

(b) (i) Secondly \mathbf{A} is defined from its vortices

$$\text{curl } \mathbf{A} = \mathbf{P} \quad \mathbf{A} \times \mathbf{n} = \mathbf{Q}, \quad (6)$$

where \mathbf{P} is a prescribed vector function of position, and \mathbf{Q} is the prescribed tangential component of \mathbf{A} on the boundary; we must have, for consistency,

$$\iiint \mathbf{P} d\tau = - \sum_{i=1}^p \iint \mathbf{Q} dS_i.$$

The necessary and sufficient condition for I to have a minimum is now

$$\text{div}(\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) = 0. \quad (7)$$

(ii) Again \mathbf{A} may not have its vortices defined so rigidly, but have

$$\text{curl } \mathbf{A} = \mathbf{P} \quad \text{and} \quad \iint (\mathbf{A} \times \mathbf{n}) dS_i = \mathbf{R}_i \quad (i=1, 2, \dots, p). \quad . . . (8)$$

\mathbf{R}_i is a vector associated with the bounding surface S_i .

The necessary and sufficient conditions are now

$$\text{div}(\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) = 0 \quad \text{and} \quad (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) \cdot \mathbf{n} = 0, \quad . . . (9)$$

and of course
$$\iiint \mathbf{P} d\tau = - \sum_{i=1}^p \mathbf{R}_i.$$

3. PROOF OF THE THEOREM.

We have to investigate the minimum of the integral

$$I = \iiint (\mathbf{A} \cdot \boldsymbol{\varphi} \cdot \mathbf{A} + 2\mathbf{C} \cdot \mathbf{A}) d\tau.$$

If we make an arbitrary variation $\delta\mathbf{A}$ in \mathbf{A} the consequent change in I is

$$\delta I = 2 \iiint \delta\mathbf{A} \cdot (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) d\tau + \iiint \delta\mathbf{A} \cdot \boldsymbol{\varphi} \cdot \delta\mathbf{A} d\tau. \quad . . . (10)$$

Since the components of $\boldsymbol{\varphi}$ form the coefficients of a positive definite quadratic form, the necessary and sufficient condition that I have a minimum is the vanishing of the integral

$$\delta I' = \iiint \delta\mathbf{A} \cdot (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) d\tau. \quad (11)$$

The proof now falls into different parts corresponding to the enunciation.

(a) Since $\delta\mathbf{A}$ is arbitrary we can choose it to be zero everywhere except along a narrow tube surrounding a closed curve and here parallel to the element of arc of the curve at the point. Then we have

$$\iiint \delta\mathbf{A} \cdot (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) d\tau = \alpha \oint (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) \cdot d\mathbf{s},$$

where the integral on the right is a line-integral taken around the closed curve and α stands for the product of $|\delta\mathbf{A}|$ and the cross-section of the tube which product we make constant. Since the circuit is arbitrary we must have

$$\text{curl}(\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) = 0. \quad (12)$$

Hence we see that the condition is necessary.

We may now write

$$\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C} = \text{grad } u. \quad (13)$$

$$\begin{aligned} \therefore \delta I' &= \iiint \delta \mathbf{A} \cdot (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) d\tau = \iiint \delta \mathbf{A} \cdot \text{grad } u d\tau \\ &= \iiint \text{div } (u \delta \mathbf{A}) d\tau - \iiint u \text{div } \delta \mathbf{A} d\tau. \end{aligned}$$

The last integral on the r.h.s. vanishes, since $\text{div } \mathbf{A} = \text{div } (\mathbf{A} + \delta \mathbf{A}) = M$ and the first integral becomes a sum of surface integrals over the boundary (the integral at infinity vanishing).

$$\therefore \delta I' = \sum_{i=1}^p \iint u \delta \mathbf{A} \cdot \mathbf{n} dS_i.$$

Now under condition (a) (i) $\mathbf{A} \cdot \mathbf{n}$ is prescribed and hence $\delta \mathbf{A} \cdot \mathbf{n}$ vanishes;

$$\therefore \delta I' = 0. \quad (14)$$

This proves that in case (a) (i) the vanishing of $\text{curl } (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C})$ is the necessary and sufficient condition.

Under conditions (a) (ii) we have

$$\delta I' = \sum_{i=1}^p \iint u \delta \mathbf{A} \cdot \mathbf{n} dS_i = 0,$$

subject to the condition $\iint \delta \mathbf{A} \cdot \mathbf{n} dS_i = 0 \quad (i=1, 2, \dots, p)$.

We can easily see that, on each of the surfaces S_i ,

$$u = u_i = \text{constant} \quad (15)$$

is sufficient to ensure the vanishing of $\delta I'$. To prove that it is also necessary consider a choice of $\delta \mathbf{A}$ such that $\delta \mathbf{A} \cdot \mathbf{n}$ has equal and opposite values at two equal elements of S_i and is zero elsewhere. This then ensures that

$$\iint \delta \mathbf{A} \cdot \mathbf{n} dS_i = 0.$$

But with this choice of $\delta \mathbf{A}$ we get

$$\begin{aligned} \delta I' &= u_1 (\delta \mathbf{A} \cdot \mathbf{n})_1 \alpha + u_2 (\delta \mathbf{A} \cdot \mathbf{n})_2 \kappa \\ &= (u_1 - u_2) (\delta \mathbf{A} \cdot \mathbf{n}) \alpha, \end{aligned}$$

where α is the element of area at the points concerned. Hence $\delta I'$ can vanish only if $u_1 = u_2$. But the two points are quite arbitrary, hence

$$u = u_i = \text{constant} \quad (i=1, 2, \dots, p)$$

is a necessary condition for the vanishing of $\delta I'$. But $u_i = \text{constant}$ means that the surfaces S_i are level-surfaces of the scalar field u and hence the tangential components of $\text{grad } u$ must vanish. Hence, in case (a) (ii) the necessary and sufficient conditions are

$$\text{curl } (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) = 0 \text{ and } (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) \times \mathbf{n} = 0. \quad (16)$$

(b) The proof under these conditions follows very similar lines. We have to find the conditions under which

$$\delta I' = \iiint \delta \mathbf{A} \cdot (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) d\tau$$

shall vanish.

Since now $\text{curl } \mathbf{A}$ is prescribed, we choose $\delta \mathbf{A}$ to vanish everywhere except in a thin shell-like volume over an arbitrary closed surface and $\delta \mathbf{A}$ here is normal to the surface;

$$\therefore \delta I' = \beta \iint (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) \cdot \mathbf{n} dS,$$

where β is the product of $|\delta \mathbf{A}|$ with the normal thickness of the shell and the integral is taken over this closed surface;

$$\therefore \delta I' = \beta \iiint \text{div} (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) d\tau,$$

taken throughout the volume enclosed by the surface. Since this is arbitrary,

$$\text{div} (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) = 0 \quad (17)$$

is a necessary condition for $\delta I'$ to vanish.

Hence we may write

$$\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C} = \text{curl } \mathbf{F}; \quad (18)$$

$$\begin{aligned} \therefore \delta I' &= \iiint \delta \mathbf{A} \cdot \text{curl } \mathbf{F} d\tau \\ &= \iiint \text{div} (\mathbf{F} \times \delta \mathbf{A}) d\tau + \iiint \mathbf{F} \cdot \text{curl } \delta \mathbf{A} d\tau \\ &= \sum_{i=1}^p \iint (\mathbf{F} \times \delta \mathbf{A} \cdot \mathbf{n}) dS_i, \end{aligned}$$

since $\text{curl } \delta \mathbf{A}$ vanishes everywhere.

Now under conditions (b) (i) $\delta \mathbf{A} \cdot \mathbf{n} = 0$;

$$\therefore \delta I' = 0, \quad (19)$$

and we see that in case (b) (i) the necessary and sufficient condition for I to have a minimum is

$$\text{div} (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) = 0.$$

Under conditions (b) (ii) we have, in order that $\delta I'$ shall vanish,

$$\sum_{i=1}^p \iint (\mathbf{F} \cdot \delta \mathbf{A} \times \mathbf{n}) dS_i = 0,$$

subject to the condition that

$$\iint (\delta \mathbf{A} \times \mathbf{n}) dS_i = 0 \quad (i = 1, 2, \dots, p).$$

Clearly $\mathbf{F} = \mathbf{F}_i = \text{constant vector}$ over each surface S_i is sufficient to satisfy the condition and, as in (a) (ii), by choosing a special form for $\delta \mathbf{A}$ we can show that it is also necessary. Now if

$$\mathbf{F} = \mathbf{F}_i = \text{constant vector} \quad (20)$$

over each bounding surface, the line integral of \mathbf{F} around any closed curve in the surface must vanish. This implies therefore that

$$(\text{curl } \mathbf{F}) \cdot \mathbf{n} = 0,$$

$$i. e. \quad (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) \cdot \mathbf{n} = 0. \quad . \quad . \quad . \quad . \quad (21)$$

Hence, in case (b) (ii) the necessary and sufficient conditions to ensure a minimum for I are

$$\text{div } (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) = 0 \quad \text{and} \quad (\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}) \cdot \mathbf{n} = 0. \quad . \quad . \quad . \quad (22)$$

4. APPLICATIONS.

(a) Thomson's theorem. If we put

$$\mathbf{C} = 0 \quad \text{and} \quad \boldsymbol{\varphi} = (\mathbf{ii} + \mathbf{jj} + \mathbf{kk})/\epsilon$$

we can identify \mathbf{A} as the electric displacement \mathbf{D} defined in terms of its sources

$$\text{div } \mathbf{D} = 4\pi\rho,$$

and the electric field \mathbf{E} is the vector $\boldsymbol{\varphi} \cdot \mathbf{A} + \mathbf{C}$ for

$$\mathbf{E} = \mathbf{D}/\epsilon = \boldsymbol{\varphi} \cdot \mathbf{D},$$

and $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are unit vectors along the coordinate axes which form the isotropic dyadic $\boldsymbol{\varphi}$ here. The boundary conditions imposed on \mathbf{D} are that the charge q_i on the conductor S_i is fixed,

$$4\pi q_i = \iint \mathbf{D} \cdot \mathbf{n} \, dS_i.$$

Then the conditions that the energy

$$8\pi U = \iiint \mathbf{A} \cdot \boldsymbol{\varphi} \cdot \mathbf{A} \, d\tau = \iiint \mathbf{D} \cdot \mathbf{E} \, d\tau = \iiint \epsilon \mathbf{E}^2 \, d\tau \quad . \quad . \quad (24)$$

shall be a minimum are

$$\text{curl } \mathbf{E} = 0 \quad \text{and} \quad \mathbf{E} \times \mathbf{n} = 0.$$

The latter is equivalent to the condition that the conductors are equipotential surfaces. The theorem is seen to hold, in addition, for a non-isotropic dielectric which is specified by a "dielectric tensor" ϵ ; in this case we identify $\boldsymbol{\varphi}$ with the inverse tensor ϵ^{-1} .

(b) Minimum heat theorem. The current density vector is given by

$$\mathbf{J} = \sigma(\mathbf{E} + \mathbf{E}'),$$

and the divergence of \mathbf{J} is specified together with its normal component on the boundary. If we write

$$\mathbf{E} = \mathbf{J}/\sigma - \mathbf{E}'$$

we can identify

$$\boldsymbol{\varphi} = (\mathbf{ii} + \mathbf{jj} + \mathbf{kk})/\sigma$$

and

$$\mathbf{C} = -\mathbf{E}'.$$

Then, in an isotropic conductor

$$H = \iiint (\mathbf{J}^2/\sigma - 2\mathbf{J} \cdot \mathbf{E}') d\tau$$

has a minimum value when curl \mathbf{E} vanishes, *i. e.* in the actual distribution. Once again the theorem is true for a non-isotropic conductor if a "resistivity tensor" is used in place of the φ used here.

(c) The most interesting application is that to the magnetic field, which does not seem to have been made before.

Consider a medium which is linear, but which need not be isotropic, and introduce the magnetic induction

$$\mathbf{B} = \mu\mathbf{H} + 4\pi\mu_0\mathbf{M}, \quad (25)$$

where μ is the permeability of the medium, μ_0 that of a vacuum and \mathbf{M} represents the permanent magnetization. The magnetic field is determined from its vortices

$$\text{curl } \mathbf{H} = 4\pi\mathbf{J} \quad \text{and} \quad \mathbf{H} \times \mathbf{n} = 4\pi\mathbf{K}. \quad (26)$$

where \mathbf{J} is the current density, and \mathbf{K} the vector representing the surface densities of the currents in the boundary conductors. Then the integral which is minimized when

$$\text{div } \mathbf{B} = 0$$

is

$$\begin{aligned} 8\pi W &= \iiint (\mu\mathbf{H}^2 + 8\pi\mu_0\mathbf{M} \cdot \mathbf{H}) d\tau \\ &= 2\iiint \mathbf{H} \cdot (\mu\mathbf{H} + 4\pi\mu_0\mathbf{M}) d\tau - \iiint \mu\mathbf{H}^2 d\tau; \end{aligned}$$

$$\therefore W = \frac{1}{4\pi} \iiint \mathbf{B} \cdot \mathbf{H} d\tau - \frac{1}{8\pi} \iiint \mu\mathbf{H}^2 d\tau.$$

In the interpretation of Thomson's theorem the energy U of equation (24), which is minimized, is also used as a force-function to determine the forces between conductors etc. by differentiation. The value here obtained for the magnetic energy agrees with Livens's⁽³⁾ value for the magnetic force-function instead of the density $\mu\mathbf{H}^2/8\pi$ usually adopted, but which leads to some confusion with signs.

SUMMARY.

A unification of the minimum energy theorems encountered in electromagnetic theory has been stated, and the extension of the theorem to non-isotropic dielectrics and conductors has been made. An application of the theorem to the magnetic field suggests a modified form for the density of the magnetic force-function which agrees with Livens's expression for this quantity.

REFERENCES.

- GIBBS, J. W., *Collected Works*, Vol. II., Pt. 2, p. 42.
 LIVENS, G. H., 1945, *Phil. Mag.*, **36**, 1; 1947, *Ibid.*, **38**, 543.
 THOMSON, W., *Papers on Electrostatics and Magnetism*, XIII.

XXI. *The Focusing of β -rays in a Prolate Spheroidal Magnetic Field.*

By H. O. W. RICHARDSON,

Department of Natural Philosophy, The University, Edinburgh*.

[Received September 29, 1948.]

β -RAY spectrometers based on the axially symmetric magnetic lens can be made to have a large solid angle of collection which is of value in the study of weak or short-lived sources of β - and γ -radiation. It is thus of interest to survey theoretically various forms of axially symmetric field in order to choose that of highest resolving power consistent with high intensity.

In describing the action of a magnetic lens we may make a distinction between axial and meridional focusing. The former occurs because rays emitted from a point on the axis of symmetry with the same momentum and the same inclination α to the axis must all intersect the axis again at a common point.

If we use cylindrical coordinates y , z , and ϕ to denote radial, axial, and azimuthal displacements, we can describe the spiral path of each ray by y , z coordinates in a meridional plane making an angle ϕ with the plane of emission and rotating as ϕ varies. If the rotation of the image does not matter ϕ can be ignored and all rays starting from the axis can be represented in the same meridional plane. The second or meridional kind of focusing arises from the intersection in this representational plane of rays emitted with different inclinations α . The intersection occurs on or near the caustic curve or envelope of the rays at a point not in general lying on the axis. In the usual types of lens the meridional and axial intersections only coincide for paraxial rays having zero angle of emission, which is too small for use in β -ray spectroscopy.

A spectrometer using meridional focusing has been described by Witcher (1941) in which the particles are collected through an annular slit which selects a part of the caustic. Most other lens spectrometers seem to use axial focusing.

In order to get improved resolving power it would seem desirable to work with larger angles of emission than have commonly been used and, if possible, to retain the advantage of the low background effect possessed by a small final collecting chamber.

We discuss the focusing properties of a field in which the lines of force are confocal ellipses and the equipotential surfaces are hyperboloids of revolution of two sheets. The rays in the meridional plane have the

* Communicated by Prof. N. Feather.

property of crossing over at two successive sites of which the first lies in the region of largest radial displacement from the axis and is of the same nature as the cross-over selected by the collector in the spectrometer of Witcher. The second cross-over is conveniently near the axis and is well suited for detection by a small Geiger-Müller tube with a narrow ring-shaped window encircling the cathode cylinder.

The meridional focusing bears some resemblance to the ordinary focusing of a concave mirror having chromatic aberration. This resemblance can serve as a guide in finding the focusing conditions, the second cross-over being the analogue of a real inverted image of about unit magnification.

1. THE VECTOR POTENTIAL AND THE MAGNETO-MOTIVE FORCE.

It is well known that in the case of an axially symmetric field the path of an electron in the rotating meridional (y, z) plane is governed by a scalar potential energy $V(y, z)$. The vector potential \mathbf{A} of the magnetic field reduces to the azimuthal ϕ component alone which is normal to the meridional plane and may be written $A(y, z)$. When the electron path intersects the symmetry axis the potential V is proportional to A^2 .

Let us put the origin of the cylindrical coordinates in the equatorial or mid-plane half-way between the two foci of the conicoids which are $2a$ apart at $y=0, z=a$ and $y=0, z=-a$. Introduce prolate spheroidal coordinates ξ and η satisfying the equations

$$\xi = (r_1 - r_2)/2a \quad \text{and} \quad \eta = (r_1 + r_2)/2a$$

where r_1 and r_2 are the distances of a point (ξ, η) from the foci at $z=-a$ and $z=+a$, respectively. Then the equation of a line of force is $\eta = \text{constant}$, $\phi = \text{constant}$, while that of a magnetic equipotential surface is $\xi = \text{constant}$.

The connection between the two systems of coordinates is given by

$$y = a\sqrt{(1-\xi^2)(\eta^2-1)}; \quad z = a\eta\xi.$$

The elements of length ds_1 and ds_2 , measured respectively along and at right angles to a line of force, in the meridional plane, are given by

$$ds_1 = a \cdot \sqrt{(\eta^2 - \xi^2)/(1 - \xi^2)} \cdot d\xi,$$

$$ds_2 = a \cdot \sqrt{(\eta^2 - \xi^2)/(\eta^2 - 1)} \cdot d\eta.$$

Consider a tube of force bounded by the coordinate surfaces η and $\eta + d\eta$, ϕ and $\phi + d\phi$. Let $d\sigma$ be the area of cross-section of this tube at any point and dN be the (constant) flux which it carries. Then H , the magnetic intensity along the tube, is given by $H = dN/d\sigma$, or, since

$$d\sigma = y \, d\phi \cdot ds_2 = a^2 \sqrt{(1-\xi^2)(\eta^2-\xi^2)} \cdot d\phi \, d\eta,$$

$$H = dN/[a^2 \cdot d\phi \, d\eta \cdot \sqrt{(1-\xi^2)(\eta^2-\xi^2)}]. \quad \dots \dots (1)$$

Let M be the magneto-motive potential difference between any two equipotential surfaces $\xi = \text{constant}$. Then

$$M = \int_{\xi_1}^{\xi_2} \mathbf{H} \cdot d\mathbf{s}_1 = \int_{\xi_1}^{\xi_2} \frac{dN}{a^2 \cdot d\phi d\eta} \cdot \frac{ad\xi}{(1-\xi^2)} \quad \dots \quad (2)$$

The quantity $dN/a^2 d\phi d\eta$ is constant along a given tube of force by our geometrical definition of a tube. From equation (2) we see that it must also be constant for all values of η and ξ because M must be independent of the particular tube chosen for the path of integration between the two equipotential surfaces.

To find the value of the constant, put $\eta=1$, $\xi=0$ in (1), thus obtaining H_0 , the magnetic intensity at the central point of the mid-plane. We get $H_0 = dN/a^2 d\phi d\eta$ which is thus the constant in question and we have

$$H = H_0 / \sqrt{(1-\xi^2)(\eta^2-\xi^2)}, \quad \dots \quad (3)$$

$$M = H_0 a \int_{\xi_1}^{\xi_2} d\xi / (1-\xi^2) = H_0 a (\tanh^{-1} \xi_2 - \tanh^{-1} \xi_1) \quad \dots \quad (4)$$

The vector potential \mathbf{A} , which satisfies $\text{curl } \mathbf{A} = \mathbf{H}$, can be found if we know N , the magnetic flux enclosed by a circle of radius y drawn in any plane normal to the axis of symmetry $y=0$. By Stokes' theorem, integrating round the circle

$$N = \int (\mathbf{H} \cdot d\boldsymbol{\sigma}) = \oint (\mathbf{A} \cdot d\mathbf{s}) = 2\pi y A.$$

For a circle in the mid-plane ($\xi=0$) we have from (3) that $H = H_m = H_0/\eta$ and

$$N = N_m = \int_0^y 2\pi y H_m dy = \int_1^\eta 2\pi a^2 H_0 d\eta = 2\pi a^2 H_0 (\eta - 1) \quad \dots \quad (5)$$

No flux crosses the spheroidal surfaces $\eta = \text{constant}$ so that the corresponding ellipses in the meridional plane are lines of equal flux N given by (5). (5) thus applies also to points not in the mid-plane and is quite general. The vector potential $N/2\pi y$ is therefore

$$A = \frac{a^2 H_0 (\eta - 1)}{y} = \frac{a H_0 (\eta - 1)}{\sqrt{(1-\xi^2)(\eta^2-1)}} \quad \dots \quad (6)$$

With fixed H_0 , the contours of equal potential energy V in the meridional plane can be plotted from (6) if the ellipses $\eta = \text{constant}$ are first drawn. It is more convenient to express (6) in polar coordinates r_1, θ , choosing $\theta=0$ to point along the axis towards the right, with origin at $y=0, z=-a$. This gives the equation of the contours $A = \text{constant}$.

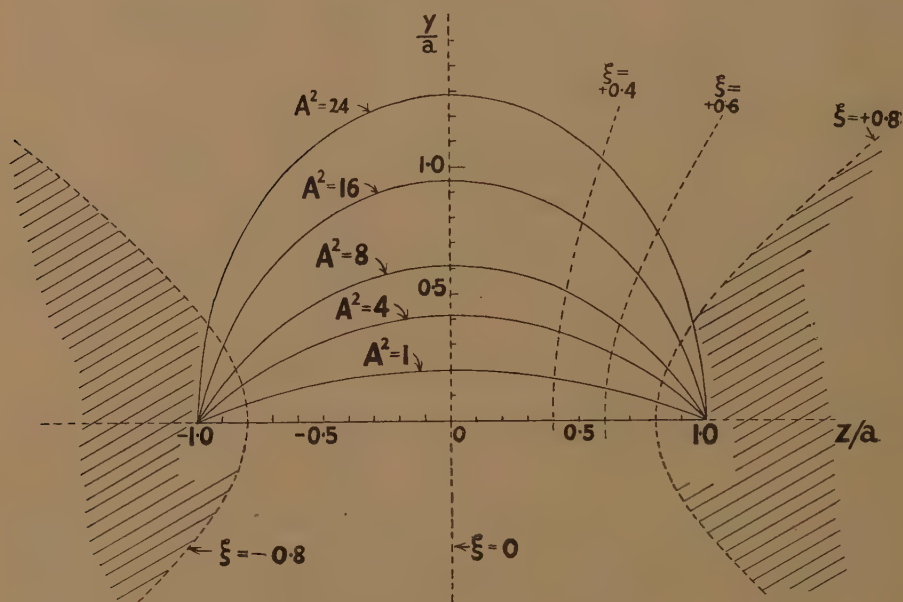
$$\frac{r_1}{a} = \frac{1 - \cos \theta - 2b \sin \theta}{b \sin \theta (b \sin \theta - 1)} \quad \dots \quad (7)$$

where $b = a(\eta - 1)/y = A/aH_0$.

From (7), since $r_1=2a$ when $\theta=0$, for all values of b , we see that the contours all pass through the foci.

The contours, drawn in fig. 1, are nearly circular and give a bowl-shaped potential barrier having a converging action on a beam of particles reflected from it. In the mid-plane $z=0$ the potential gradient is nearly constant except near $y=0$, where it becomes zero. In the case of a uniform field, in contrast, the gradient increases linearly with y . In the mid-plane of the short magnetic lens formed by a short coil carrying a current, the gradient increases even more rapidly with y and the contours $A^2=\text{constant}$ are strongly convex towards the axis.

Fig. 1.



Contours of equal vector potential A in the meridional (y, z) plane. The hyperbolæ $\xi = \pm 0.8$ show a possible axial section of the magnet pole-faces.

The contours of A^2 for the spheroidal field bear some resemblance to the equipotentials of a two-tube electrostatic lens. The use of this field as an electron mirror has been frequently discussed, for example by Nicoll (1938).

2. THE CURVATURE OF THE REFLECTING MERIDIONAL CONTOURS.

Within the limits in which the field behaves like a cylindrical reflector, we can estimate its focal length if we know the radius of curvature of the contours responsible for the reflection.

The radius ρ_0 of the contour $A^2=\text{constant}$ at the point where it crosses the mid-plane can readily be found if either of the equations (6) or (7)

is expressed in cartesian form. Let us eliminate η between the equation $c=(\eta-1)/y=A/a^2H_0$ and the cartesian equation of the ellipse $\eta=\text{constant}$,

$$\frac{z^2}{a^2\eta^2} + \frac{y^2}{a^2(\eta^2-1)} = 1.$$

This gives for the cartesian equation of the contour,

$$(1+yc)^2(a^2-y/c(y+2))-z^2=0. \quad (8)$$

In the mid-plane $z=0$ the value of y is a maximum so that there

$$\frac{dy}{dz}=0 \quad \text{and} \quad \rho = \left| \frac{(d^2y)^{-1}}{dz^2} \right| = \rho_0.$$

Differentiating we get, along $z=0$, where $y=y_0$

$$\rho_0 = a^2c(1+y_0c) - y_0 \left(\frac{1+y_0C}{2+y_0C} \right) - \frac{1}{c} \left(\frac{1+y_0C}{2+y_0C} \right)^2. \quad (9)$$

This can be simplified by using polar coordinates with the origin, as before, at $y=0$, $z=-a$, so that $y_0=a \tan \theta$; $1+cy_0=\sec \theta$. Then

$$\rho_0 = a/\sin \theta (1 + \cos \theta) \quad (10)$$

ρ_0 has its minimum value $0.770a$ when $\theta=60^\circ$, $y_0=1.732a$. When $\theta=41^\circ$, $\rho_0=y_0=0.869a$. The centre of curvature then lies on the axis of symmetry. We therefore expect that in order to form a real image of the source on or near the axis we must use a selecting annulus in the mid-plane with a radius of about $0.9a$, where a is half the distance between the common foci of the hyperboloids and ellipses. To get a more exact value for the optimum radius of the selecting annulus we can proceed either experimentally or by the computation of electron orbits in the meridional plane by numerical integration of the equations of motion.

3. THE RELATIVISTIC EQUATIONS OF MOTION.

For motion of a β -ray in a magnetic field the energy and therefore the mass remain constant provided that the acceleration is so small that it causes negligible loss of energy by radiation.

The constancy of mass makes it allowable to use non-relativistic methods for a given value of the momentum. It is not immediately clear how to compute trajectories of different momenta because here both the masses and the velocities will be different. The work may be reduced by applying a general theorem which holds for the motion of a charged particle in any static magnetic field if other fields are absent.

Let $m=m_0/(1-v^2/c^2)^{1/2}$, where m_0 is the rest mass. The relativistic equation of motion is, since the speed v is constant,

$$\frac{d^2\mathbf{r}}{dt^2} = \frac{e}{m} \left[\frac{d\mathbf{r}}{dt} \times \mathbf{H} \right] \quad (11)$$

where e and H are in e.m.u.

Following the procedure of Störmer (1933) we put $t=s/v$ where s is the distance measured along the spiral path. Then

$$\frac{d^2\mathbf{r}}{ds^2} = \frac{e}{mv} \left[\frac{d\mathbf{r}}{ds} \times \mathbf{H} \right] = \frac{He}{p} \left[\frac{d\mathbf{r}}{ds} \times \mathbf{h} \right] \quad . \quad . \quad . \quad (12)$$

where \mathbf{h} is a unit vector in the direction of the field \mathbf{H} and $p=mv$ is the magnitude of the relativistic momentum.

Thus the shape and size of the orbits are invariant under any transformation which leaves He/mv unchanged, because the other quantities in (12) are geometrical. We may, for example, choose $m=1$ gm. for ease of computation and it is then only necessary to reduce v in a ratio which maintains the correct value of mv .

A second consequence of (12) is that to compute an orbit of higher momentum p we may retain the convenient value $m=1$ provided only that we increase v in proportion to p . A third consequence, important in magnetic spectrometry, is that to bring particles of higher momentum to a focus at a fixed point it is only necessary to increase H everywhere along the path in proportion to the increase in momentum.

The equations of motion* in an axially symmetric field are, first

$$my^2\dot{\phi}=C-eAy \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where the constant C is the generalized momentum associated with the coordinate ϕ , and

$$\left. \begin{aligned} m\ddot{y} &= -\frac{1}{2m} \frac{\partial}{\partial y} (C/y - eA)^2, \\ m\ddot{z} &= -\frac{1}{2m} \frac{\partial}{\partial z} (C/y - eA)^2. \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad . \quad (14)$$

We see that the motion in the meridional (y, z) plane is that of a particle of mass m obeying Newton's laws of motion in a conservative field of force with a scalar potential energy

$$V(y, z) = \frac{1}{2m} (C/y - eA)^2. \quad . \quad . \quad . \quad . \quad . \quad (15)$$

For a ray starting on the axis of symmetry where $A=0$, we have $C=0$ and $V=e^2A^2/2m$. At the starting point, where $V=0$, we have $v=(\dot{x}_0^2 - \dot{y}_0^2)^{1/2}$ where \dot{x}_0 and \dot{y}_0 are the initial values. To compute a ray of higher momentum we can use the same values of the meridional accelerations in (14) provided we keep the same value of m and only increase the speed v .

To find the value of the field strength H_0 at the origin needed to focus a given momentum, we can use the energy equation

$$\frac{1}{2}m(\dot{y}^2 + \dot{z}^2) = \frac{1}{2}mv^2 - e^2A^2/2m. \quad . \quad . \quad . \quad . \quad (16)$$

* These follow easily from Lagrange's equations using the relativistic Lagrangian $L=mc^2(1-\sqrt{1-v^2/c^2})+e(\mathbf{v} \cdot \mathbf{A})$.

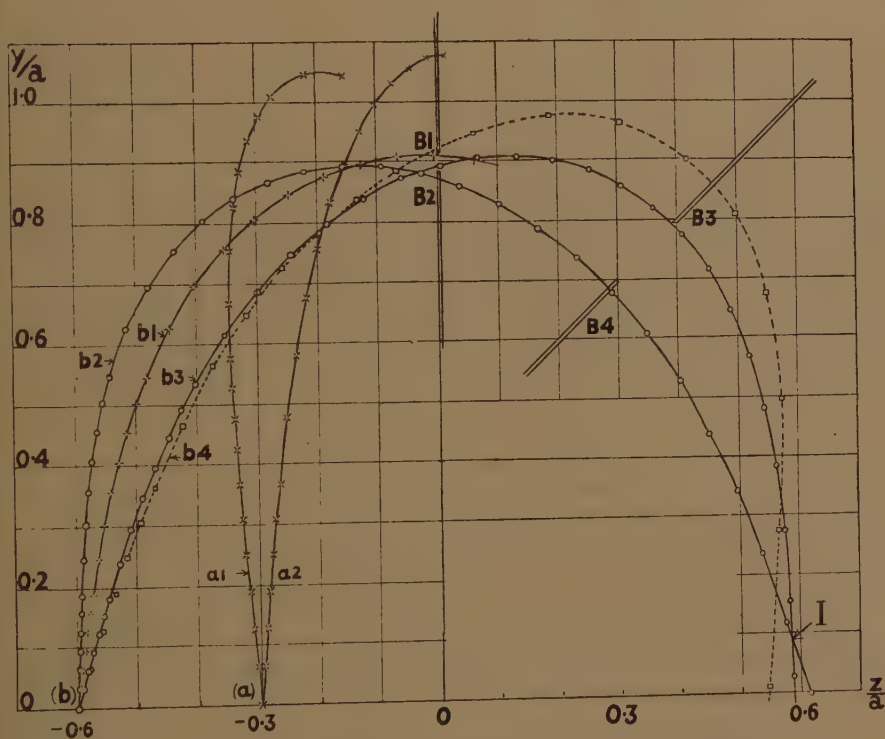
Put $H\rho = mv/e$ and substitute for A from equation (6). Then

$$\dot{y}^2 + \dot{z}^2 = v^2(1 - [aH_0/H\rho]^2 \cdot f(\eta, \xi)) \quad . \quad . \quad . \quad (17)$$

where $f(\eta, \xi)$ is the dimensionless expression $(\eta - 1)/(\eta + 1)(1 - \xi^2)$.

For the particular trajectories plotted in fig. 2 the constant $\frac{aH_0}{H\rho}$ was chosen to be $\sqrt{5}$ so that if $a = 10$ cm. we need $H_0 = 1000$ oersteds to focus rays with $H\rho = 4472$.

Fig. 2.



Computed orbits in the meridional plane. The ray b_4 , which has 5 per cent more momentum than the others, shows chromatic aberration. The points show the steps used in integration. Rays b_4 , a_1 , and I were started by using the Taylor expansion (19) only up to the 2nd derivative. They are thus less exact than b_1 , b_2 , and b_3 and may be regarded exploratory trials.

4. COMPUTATION OF THE ORBITS.

A few meridional trajectories are shown in fig. 2. They were found by step by step integration by a method differing slightly from that of Goddard (1944).

Inserting exact expressions for the derivatives of A , the equations of motion (14) become, in a form convenient for computation,

$$\left. \begin{aligned} \ddot{y} &= - \left(\frac{a^2 H_0 e}{m} \right)^2 \left(\frac{\eta-1}{y} \right) \left(\frac{1}{2a} \left[\frac{1}{r_1} + \frac{1}{r_2} \right] - \frac{\eta-1}{y^2} \right), \\ \ddot{z} &= - \left(\frac{a^2 H_0 e}{m} \right)^2 \left(\frac{\eta-1}{2ay^2} \right) \left(\frac{z+a}{r_1} - \frac{a-z}{r_2} \right). \end{aligned} \right\} \dots \quad (18)$$

In order to start the process of step by step numerical integration along an orbit, the Taylor expansion method may be used for the initial steps, each traversed in the time interval τ . Then

$$\left. \begin{aligned} y_{m+1} &= y_m + \tau \dot{y}_m + \frac{\tau^2}{2!} \ddot{y}_m + \dots \\ \dot{y}_{m+1} &= \dot{y}_m + \tau \ddot{y}_m + \frac{\tau^2}{2!} \dddot{y}_m + \dots \end{aligned} \right\} \dots \quad (19)$$

with similar expressions for z_{m+1} .

z_m, y_m are the coordinates of the electron after the m th interval τ . Our field differs from that of the short magnetic lens discussed by Goddard and Klemperer (1944) in that the intensity H at the axial source is large, although A, \ddot{z} , and \ddot{y} are zero on the axis. For this reason it was found necessary to carry (19) as far as the fourth derivatives. Manageable expressions for the higher derivatives can be found if (18) is replaced by the approximations

$$\left. \begin{aligned} \ddot{z} &= - (a^2 H_0 e / m)^2 y^2 z / 2(a^2 - z^2)^3, \\ \ddot{y} &= - (a^2 H_0 e / m)^2 y / 4(a^2 - z^2)^2. \end{aligned} \right\} \dots \quad (20)$$

These are valid near the axis $y=0$ so long as y/a may be neglected in comparison with unity.

When y_m is known for four values of m the integration is continued using the formulæ of Milne (1933)

$$y_{m+1} = y_m + y_{m-2} - y_{m-3} + \frac{\tau^2}{4} (5\ddot{y}_m + 2\ddot{y}_{m-1} + 5\ddot{y}_{m-2}) \dots \quad (21)$$

or

$$y_{m+1} = y_m + y_{m-4} - y_{m-5} + \frac{\tau^2}{48} (67\ddot{y}_m - 8\ddot{y}_{m-1} + 122\ddot{y}_{m-2} - 8\ddot{y}_{m-3} + 67\ddot{y}_{m-4}) \dots \quad (22)$$

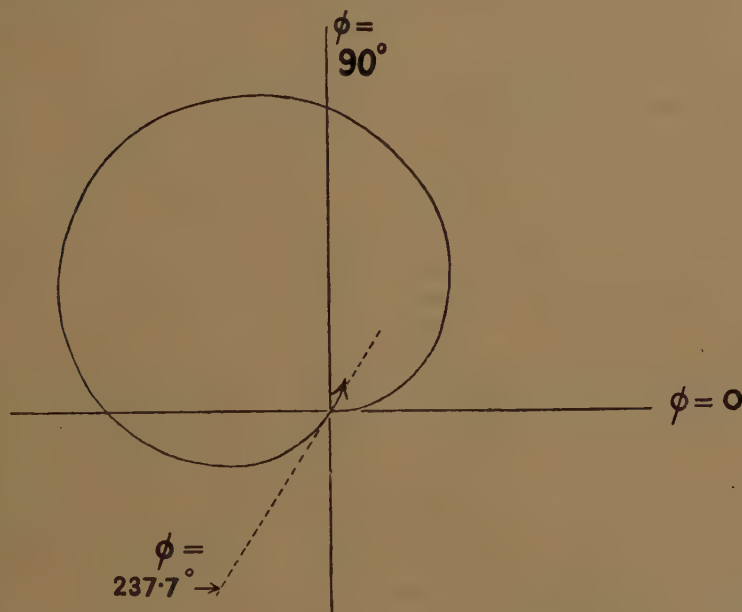
with corresponding expressions for z_{m+1} . Equation (21) is exact if fourth differences are negligible so that the initial time interval τ must be small enough to satisfy this condition. When enough values have been computed it is possible to increase the interval and to use (22) which is exact if sixth differences can be neglected.

From (13) we have

$$\phi = - \frac{e}{m} \int_0^t \frac{A}{y} dt = - \frac{ea^2}{m} H_0 \int_0^t \frac{\eta-1}{y^2} \cdot dt. \dots \quad (23)$$

Numerical integration of this for the ray b_2 in fig. 2 gives the curve in fig. 3, the projection of this ray on the (y, ϕ) plane normal to the axis of symmetry. The azimuth ϕ of the electron changes by 237.7° between its departure from and its return to the axis. This exceeds the rotation of 180° given by a uniform field because the path spirals rapidly near the axis where the field is strong.

Fig. 3.



Projection of the ray b_2 (fig. 1) on the (y, ϕ) plane normal to the axis of symmetry.

5. THE POSITION OF THE SOURCE AND OF THE DEFINING SLITS.

It can be seen from the computed paths in fig. 2 that the envelope of a family of rays of equal momentum from an axial source lies roughly parallel to the contours of equal vector potential A in the region where the first crossing over occurs. This justifies the use of the analogy with reflection at a cylindrical mirror having chromatic aberration.

This shape of envelope is quite different from that given by a uniform field H , or by the short magnetic lens computed by Goddard and Klemperer (1944), where the envelope is inclined to the contours of A and the analogy with specular reflection is weak. In the spheroidal field the comparative constancy of the peak values of A reached by the rays of a family is due to the bowl-like shape of the contours. A ray with a large angle of emission passes through a region in which the z -acceleration is large so that radial momentum is transferred from the y to the z direction. Thus the large initial radial velocity is reduced, and we may have a case in which the rays of a family cross the mid-plane with axial velocities which are nearly equalized.

Two families (a) and (b) have been computed with sources at (a) $z_0 = -0.3a$ and (b) $z_0 = -0.6a$. Let us define the ray which crosses the mid-plane normally as the central ray. Then this ray must be symmetrical about the mid-plane. The emission angle α of the central ray of family (a) is about 85° and for (b) it is about 80° . If a source is to be used which is deposited on a flat surface normal to the axis of the magnet, case (b) is to be preferred because the smaller value of α allows the escape of β -rays with less loss of energy in passing through any thin surface layer of matter present on the source.

A second advantage of (b) is that the degree of concentration of the rays where they cross the mid-plane seems to be greater than for position (a) so that an annular defining slit of given width would receive rays emitted over a larger solid angle. A position for the source and annulus must almost certainly exist which gives still greater focusing on the mid-plane. The trajectories (b) in fig. 2 show that rays emitted with α between 74° and 84° will all pass through an annulus of width $0.02a$. Thus, for a point source, 8.5 per cent of the total emission is collected by an annular slit whose width is 2.2 per cent of its radius.

A narrow defining annular slit B_1B_2 placed in the mid-plane will pass no rays of momentum less than that of the central ray which just reaches the slit. Most rays of higher momentum will also be stopped but some rays such as b_4 in fig. 2, which fall on the mid-plane with appreciable inclination, will pass. To exclude these inclined rays a subsidiary slit B_3B_4 may be used. Its selecting action will resemble that of the main defining slit in the semicircular spectrograph by reducing the tail present on the low-energy side of a β -ray line.

In section 2 we found that the centre of curvature of the contours of the vector potential A lies close to the axis $y=0$ when $y_0 = \rho_0 = 0.9a$. The rays of family (b) are reflected near this contour so that, if the analogy with the cylindrical mirror is valid, we should expect them to intersect at the second cross-over fairly close to the axis. In fact the rays b_2 and b_3 of fig. 2 do intersect near the axis at about $y=0.09a$. The exact position of intersection is rather sensitive to small errors in computation but it clearly lies very close to the symmetrical ray at the centre of the focused beam. For the trajectories of fig. 2 the accuracy of calculation of the accelerations probably did not exceed about 1 in 1000. A family such as (b), but with a slightly lower momentum, would give an image I lying exactly on the axis.

6. THE MOMENTUM DISPERSION AND RESOLVING POWER.

The resolving power will be due partly to the selecting action of the slits B_1B_2 and B_3B_4 and partly to the chromatic aberration in the position of the real image formed near the Geiger-Müller tube. Let y_m be the value of y at which a central ray of momentum p falls normally on the mid-plane. We may define the momentum dispersion D in the mid-plane as $p \, dy_m / dp$, a quantity proportional to the y -displacement caused by a given fractional change in momentum.

In the case of the semicircular spectrograph, the source-image distance is 2ρ and its dispersion is

$$D = \rho \, d(2\rho)/d\rho = 2\rho/H\epsilon = 2\rho, \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

D is thus proportional to the length of the spectrograph.

In the spheroidal field D could be found by computing values of y_m for different momenta. This would be laborious because the correct angle of emission α which gives a central ray falling normally on the mid-plane must be found by trial and error.

We can get a rough estimate of D by using the result that the y -acceleration \ddot{y} varies but little over most of the field. If we approximate by assuming it exactly constant we have the case of projectiles moving freely under gravity in parabolic orbits. Let z_0 be the z -coordinate of the source S . Then

$$D = 2y_m(4y_m^2 + z_0^2)/(4y_m^2 - z_0^2). \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

If $z_0 = -0.66y_m$, $D = 2.5y_m$. The dispersion is thus similar to that found with semicircular focusing. It seems probable that the momentum resolving power due to the annular slits B_1B_2 and B_3B_4 will be limited in practice by the finite size of the source rather than by aberration.

The chromatic aberration of the final image can be regarded as a momentum dispersion of the point of intersection I of the family of rays from a point source. This dispersion will occur along a line which passes through the collecting slit, and is nearly perpendicular to the axis $y=0$. Its value will be about $2\frac{1}{2}$ times the dispersion D in the mid-plane, because rays of higher momentum are reflected at an equipotential more distant from the axis so that (1) the distance from the object to the mirror is increased and (2) the radius of curvature of the mirror is decreased in accordance with equation (10). Rays of higher momentum such as b_4 in fig. 2 thus intersect at points in front of the collecting slit and if the slit is narrow, rays from a distant intersection will have but small chance of entry. To get a large contribution to the resolving power from this effect it is necessary that the width of the image due to the size of the source, and to non-chromatic aberrations, be very small.

7. CONCLUSION.

It seems probable that by using a prolate spheroidal field an axially symmetric wide-angle β -spectrograph may be designed with a resolving power comparable with that attainable by using semicircular focusing in a uniform field.

Two types of source could be used. For the investigation of β -ray lines a very fine activated wire might be mounted along the axis. In this case the counter slit must be wide and to get high resolution one could use a narrow annular slit in the mid-plane not much wider than the thickness of the wire. This arrangement would demand high axial symmetry in the magnetic field in order to realize the full resolving power.

The second type of source would be mounted on a thin film with its plane normal to the axis. Such a disk source would usually have an appreciable diameter which would limit the resolving power due to an annulus in the mid-plane. On the other hand, the real image of such a flat source is probably very narrow, so that some additional resolution may be obtainable from the chromatic aberration of the final image.

Apparent disadvantages of the spheroidal field are that the area of the pole-faces of the magnet must be considerably larger than for a semi-circular spectrograph of similar dispersion, both because the focusing system is axially symmetric, and because there must be considerable "waste" flux passing at radii greater than that of the defining annulus. This is inevitable if the focusing field is to be approximately spheroidal in form. Again the distance between the poles is larger than with the conventional spectrograph and thus on two counts a greater expenditure of magnetizing power is needed in order to provide the necessary magnetomotive force.

I wish to thank Professor O. R. Frisch and Dr. J. C. P. Miller for helpful discussions, and Professor N. Feather and Dr. R. Schlapp for many clarifications and improvements in the paper.

SUMMARY.

A discussion is given of the applicability to β -ray spectroscopy of the magnetic field between pole-faces which are hyperboloids of revolution. Trajectories of particles emitted from points on the axis of symmetry are computed by numerical integration of the relativistic equations of motion. The rays have the property of crossing over at two successive sites in the meridional plane of which the second could be arranged to lie close to the axis. The attainable resolving power seems to be similar to that of the semicircular spectrograph, while the solid angle of collection can be made as large as 8.5 per cent of 4π .

REFERENCES.

- GODDARD, L. S., 1944, *Proc. Phys. Soc.*, **56**, 372-7.
 GODDARD, L. S., and KLEMPERER, O., 1944, *Ibid.*, **56**, 378-396.
 MILNE, W. E., 1933, *Am. Math. Mo.*, **40**, 322-327.
 NICOLL, F. H., 1938, *Proc. Phys. Soc.*, **50**, 888-898.
 STÖRMER, C., 1933, *Annalen der Phys.*, **XVI**, 685-96.
 WITCHER, C. M., 1941, *Phys. Rev.*, **60**, 32-42.

XXII. *Notices of New Books and Periodicals received.*

An Introduction to the Algebraic Geometry of a Plane. By J. W. ARCHBOLD.
[Pp. xiii+300.] (Edward Arnold & Co., London, 1948.) Price 25s. 0d.

THIS well written volume is devoted mostly to projective geometry, the metrical geometry treated being regarded as a special case. It commences with a chapter on foundations, and this includes a section on groups and fields. Chapter II. treats projective transformations, and some of the topics discussed are cross-ratios, homogeneous coordinates, Desargues' theorem and Pappus' theorem. The projective theory of conics is presented in Chapter III. and the metrical theory in Chapter IV. Various additional topics about conics, *e. g.* harmonic conic, confocals, outpolar and inpolar conics are studied in chapter V. For the first time, in an elementary book (in English), the theory of collineations and correlations in the plane is given; and there is a short section on projective invariants. Rational curves are studied in the last chapter. Multiple correspondences, and in particular (2, 2) correspondences are discussed. There is a valuable set of 125 miscellaneous examples, taken mostly from examination papers set by the University of London.

From this brief survey of the contents it will be seen that the author's approach to his subject is much broader than has been traditional in Britain, at least outside Cambridge. The aim has been to prepare the reader for the more extensive works on the same subject, and it is only fair to say that this necessary groundwork has been very well covered. Only on the last page is one introduced to the notion of a birational transformation. But the reader who has mastered what precedes will be well prepared for an understanding of the literature dealing with birational geometry. L.S.G.

Rapports et Discussions sur les Isotopes. Institut Internationale de Chimie, Solvay. [Pp. 411.] (Brussels, R. Stoops, 1948.)

THE present report of a meeting of the Solvay Congress in September, 1947, contains several articles of interest. The report on artificial isotopes contains a good account of the transuranic elements, with tables of the radioactive families with their properties. Much new material on this subject has appeared since the report was printed. An article on mass-spectrum analysis describes the apparatus and has some good reproductions of close doublets, *e. g.* of ^{12}CH and ^{13}C . There is also a discussion of the relative abundance of isotopes. Two very long articles on isotopes and band spectra of organic compounds are rather off the main line of interest. Radioactive tracer elements and their applications are dealt with in several reports, the most interesting detail in this department being perhaps the fact, which seems now to be well established, that the oxygen in photosynthesis comes from the water and not from the carbon dioxide, the primary photochemical reaction being apparently the liberation of active hydrogen from water which then reduces carbon dioxide. Many details of this process are still unknown, but a very old problem now seems on the way to a final solution. Some interesting biochemical work with radiophosphorus is also reported. The material in this volume should be of considerable interest to workers in many different fields, and it is very useful to have it collected and systematized in such a convenient form. The printing and paper are very good. and the book is strongly bound. J. R. PARTINGTON.

Vectorial Mechanics. By E. A. MILNE. [Pp. xiii+382.] (Methuen & Co., Ltd., London, 1948.) Price 36s. 0d.

It has taken a long time for vector methods to become accepted, but by now it may be safely said that the prejudice of Mathematicians brought up exclusively on systems of coordinates has been overcome. Thus the time is appropriate for the appearance of the volume under review; and, although, to judge from Professor Milne's preface, the book was planned so long ago as 1926, it is good that publication has been delayed, since this has enabled a better book to appear at a time when it will be more appreciated.

This purpose of the work is, briefly, to demonstrate the power of the operation of Vector Multiplication, $\mathbf{A}\mathbf{B}$, and the breadth of its applications. For this reason a whole chapter (21 pages) is devoted to the vector product. Some interesting examples appear here, *e.g.* the equations, $\mathbf{X}\mathbf{A}=\mathbf{B}$, $\mathbf{X}\mathbf{C}=\alpha$ have a unique solution, $\mathbf{X}=(\alpha\mathbf{A}-\mathbf{B}\mathbf{C})/(\mathbf{A}\mathbf{C})$, if $\mathbf{A}\mathbf{C}\neq 0$, but if $\mathbf{A}\mathbf{C}=0$ the general solution is $\mathbf{X}=(\mathbf{A}\mathbf{B})/\mathbf{A}^2+\lambda\mathbf{A}$ (subject to the consistency condition, $\alpha\mathbf{A}^2=\mathbf{A}\mathbf{B}\mathbf{C}$) where λ is arbitrary. A chapter is given to elementary tensor analysis and another to integral theorems of the vector calculus. Then systems of line vectors and applications to statics are considered. More than half of the book is given to the formulation of dynamics by a purely vector method. This is very thorough and is the most complete fusion of vectors and dynamics that has yet appeared. It is not merely a question of presenting the principal results of Dynamics in a vector notation. It is shown, by numerous worked examples, that problems originally or usually solved by Cartesian methods may be solved, with considerable advantage, by vector methods. The dynamics of systems of particles and of rigid bodies are considered. Gyroscopic motion and impulsive motion are also treated. The book concludes with a set of examples taken from papers set at the Imperial College of Science, London.

The present work is intended for second or third year honours students at a University; it should appeal widely and deserves a place in the library of the applied mathematician who is satisfied only with the best methods. L.S.G.

The Crystalline State.—Volume II. *The Optical Principles of the Diffraction of X-Rays.* By Prof. R. W. JAMES. (London: G. Bell & Sons Ltd.)

THE second volume of *The Crystalline State* by Professor R. W. James will receive a warm welcome from all those, including non-specialists, who are concerned with X-ray diffraction phenomena. The theoretical presentation of the optical principles, according to which diffraction patterns must be interpreted, is admirably clear, critical and well balanced. In each field, the formal mathematical derivations are given, the classical approach leading up to the wave-mechanical one where both treatments are included, and the physical significance of the results is discussed. Vector notation is used throughout and the geometrical interpretations given with reference to the reciprocal lattice, an elegant medium for exposition whose power and beauty are clearly portrayed by the author. He shows how it provides the natural physical approach to the consideration of most current X-ray diffraction problems and permits all the information derived from the study of X-ray spectra to be condensed into a vivid representation. Reference to specific experimental work is made only where it is necessary to illustrate the optical principles or to establish the applicability of the theoretical treatment.

Chapters I. to V. are devoted to the development of the quantitative theory of the diffraction of X-rays by crystals and deal with the geometrical theory intensities of reflections, the atomic scattering factor, anomalous scattering and dispersion, and the influence of temperature on diffraction phenomena.

The experimental application of the results deduced in these chapters to the intensity measurements is discussed in Chapter VI. The methods involving the use of Fourier series are introduced in Chapter VII. and Laue's development of Ewald's dynamical theory is given in Chapter VIII., which includes a discussion of Kossel lines and Mrs. Lonsdale's divergent beam photography. Chapters IX. and X. treat the scattering of X-rays by gases, liquids, amorphous and fibrous solids, emphasizing that from the point of view of the optical principles involved, there is a close relationship between the diffraction effects obtained with these materials.

The Vibrations of elastic Systems in a Sound-field. By T. VOGEL. (Published by Service de documentation et d'information technique du ministre de l'air.)

THIS paper forms one (No. 209) of a series of "Publications Scientifiques et Techniques du Ministère de L'Air" (Paris, 1948).

In the introduction the author points out that the work is primarily part of an investigation to clear up the extremely complex question of the acoustics of an enclosed body in a sound-field, relating in particular to the fuselage of an aircraft in flight.

The paper is divided into two parts. The first part deals with the theory of the transparency of elastic systems in a sound-field as applied to the case of multiple reflections and transmission at the walls of an enclosure. Expressions are derived for the sound transmitted, reflected and absorbed at a plane wall in the form of a rectangular sheet, for sound incident normally and at oblique angles. The theory of a double wall with air separation is treated as a coupled vibrating system and expressions for transparency derived as a function of frequency of the incident sound.

In the second part of the paper, a description is given of an experimental investigation of the characteristics of transparency of elastic panels. These were built into a wall, as rigidly as possible, separating two "an-echoic" rooms. Methods are described to obtain a logarithmic scale of frequencies and logarithmic amplification of the received sound—oscillographic records of these are shown in illustration. As a check on the reliability of the method of measurement, the law of decrease of intensity as a function of distance was verified in a non-echo (anechoic) room, using frequencies from 1000 to 8000 \sim sec approximately. Experiments are described which provide general verification of author's theories of transmission and reflection at single and double walls.

In conclusion the author discusses the value of the method and refers to a memoir by Dr. N. Fleming which summarizes the results of 20 years' experience at the National Physical Laboratory, Teddington. A.B.W.

Ionospheric Radio Propagation. National Bureau of Standards. [Pp. iv + 209.] (U.S. Government Printing Office, Washington, 1948.) Price \$1.

THE principal characteristics of the ionosphere and the phenomena of short-wave propagation were well known many years before the war, but the ionosphere had been regularly studied at only very few observation stations. During the war it became of great importance to be able to predict as accurately as possible the probable behaviour of a short-wave channel of communication. Consequently, a large number of observing stations were set up and many workers were employed in collating the data obtained and improving the methods of prediction. The various British and American organizations worked in close collaboration.

The book under review gives a very clear account (in rationalised M.K.S. units) of the fundamental principles involved when an E-M wave passes through an ionised medium in the presence of a magnetic field. This is followed by a discussion of methods for measuring the properties of the ionosphere and an account of its structure. Methods for predicting the maximum usable frequency and the absorption in the ionosphere are then dealt with. There is a valuable chapter dealing with the noise level to be expected in various parts of the world and the proportionate field strengths required for different types of communication.

The remainder of the book deals with practical methods for utilizing the mass of data available, in the form of charts, to decide what frequency to use for a given communication and what radiated power will be necessary. A large number of charts are given to illustrate the methods; but charts based on the latest observations are obtainable from Washington.

There are numerous references at the end of each chapter and many of these are to the important British contributions on this subject. The paper, printing and diagrams are all good. It seems rather a pity that a book of this size, having lasting value, should have only a paper cover.

C.R.S.

MARCONI

TEST GEAR EXHIBITION

Marconi Instruments Ltd. are staging a display of their current range of Communications Test Gear at their London Showrooms, 109 Eaton Square, London, S.W.1. The Exhibition will be open for two weeks, February 21st to March 4th (Monday to Friday) and there will be many working demonstrations.

Radio and Communications Engineers are cordially invited to apply for admission tickets, either to Marconi Instruments Ltd., St. Albans, Herts., or the London Showrooms.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XXIII. *The Properties of Silver Halides containing Traces of Silver Sulphide.*

By J. W. MITCHELL, D.Phil.,
H. H. Wills Physical Laboratory, University of Bristol*.

[Received December 15, 1948.]

1. INTRODUCTION.

THE purpose of this paper is to discuss the properties of silver halide crystals containing traces of silver sulphide. The subject is one of great importance because of the formation of silver sulphide during the after-ripening of photographic emulsions, the process by which high sensitivity is obtained (Mees 1942). Sheppard (1925) showed that this increase in sensitivity depended on the presence in the gelatine of organic compounds containing labile sulphur. He suggested that the sulphur might be instrumental in forming on the surface of the grains the so-called "sensitivity specks" which act as nuclei on which silver forms during photolytic action (Mees 1942, fig. 47, Sheppard *et al.* 1927). This idea was the basis of the theory of latent image formation put forward by Gurney and Mott (1938, see also Mott 1941, 1946). It was suggested that the sensitivity specks were minute crystals of silver sulphide.

In this paper we challenge this last proposition and put forward an alternative one. We base our ideas on a series of experiments on massive halide crystals containing traces of sulphide carried out by Stasiw and Teltow (1941-48). These authors have given an interpretation of their results which we believe to be incorrect and we suggest an alternative interpretation.

Our main conclusions are :

(a) At high temperatures silver halides dissolve silver sulphide, which goes into solution as a singly charged S^- ion and an F centre (anion site at which the halide ion is replaced by an electron). This is the reason why the addition of silver sulphide does not enhance the ionic conductivity as does cadmium bromide for example.

(b) Unless the crystal is very rapidly quenched, plate-like aggregates of F centres are formed which, if big enough, break away from the silver halide matrix and form colloidal silver (compare Cu_2Al in aluminium-based alloys).

(c) These aggregates act as electron traps and are the "sensitivity specks" of Sheppard's theory. In the photographic emulsion they are formed at lower temperatures by the aggregation of F centres produced during the after-ripening process.

* Communicated by the Author.

(d) Under illumination electrons leave the F centres and are transferred to the aggregates, which then increase in size through processes involving ionic migration.

(e) These processes may be of two types :

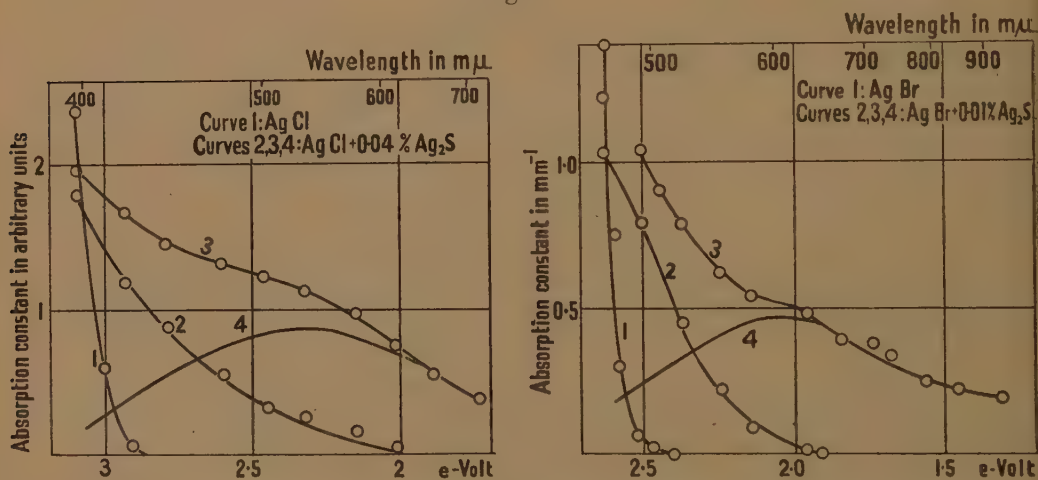
(i) A transport of cations as described by Gurney and Mott ; this will set up severe internal strains in forming a silver speck, and is usually possible only on the surface, where the silver can be pushed out as in development.

(ii) The motion towards the centre of vacant anion sites, a process by which internal image may be formed. These may be the sites vacated by electrons or, if Schottky defects exist, as assumed here, they will be present even in pure crystals.

2. THE EXPERIMENTAL RESULTS OF INVESTIGATIONS WITH MIXED CRYSTALS OF SILVER HALIDES WITH SILVER SULPHIDE.

The experimental results of Stasiw and Teltow will be summarized in this section. They prepared the mixed crystals from melts of the pure

Fig. 1.



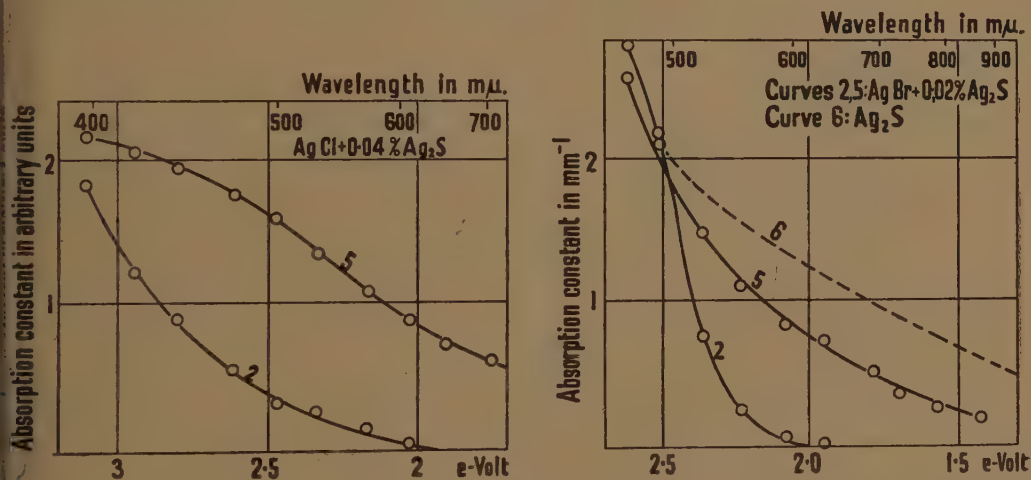
The photochemistry of silver chloride and silver bromide containing traces of silver sulphide. Curve 1. The absorption band of the pure material. Curve 2. The long wave extension of the absorption band in the presence of the impurity. Curve 3. The absorption band after irradiation. Curve 4. The absorption band assumed to be due to the presence of colloidal silver.

silver halides to which a few parts in ten thousand of silver sulphide had been added. The melt was usually transformed into a single crystal which was then quenched from a temperature a few degrees below the melting point. For the experiments, sheets with a thickness between 0.1 and 1.0 mm. were prepared by pressing the crystals out between glass plates.

The mixed crystals with silver chloride were a clear yellow, and with silver bromide a deep brown. The colour was due to an extension of the absorption region of the chloride crystals from 4400 Å to near 6000 Å and of the bromide crystals from about 5500 Å to beyond 6200 Å. This long wave absorption was associated with a new band, whose maximum appeared to lie on the short wave side of the edge of the absorption band characteristic of the lattice. This is shown in fig. 1.

After annealing at 400° C., followed by slow cooling, the silver chloride crystals became deeply coloured. Clear yellow crystals were again produced by reheating the annealed crystals to 400° C. for several hours and then quenching rapidly. Similar results were obtained with silver bromide. The absorption characteristics of the annealed crystals are shown in fig. 2.

Fig. 2.

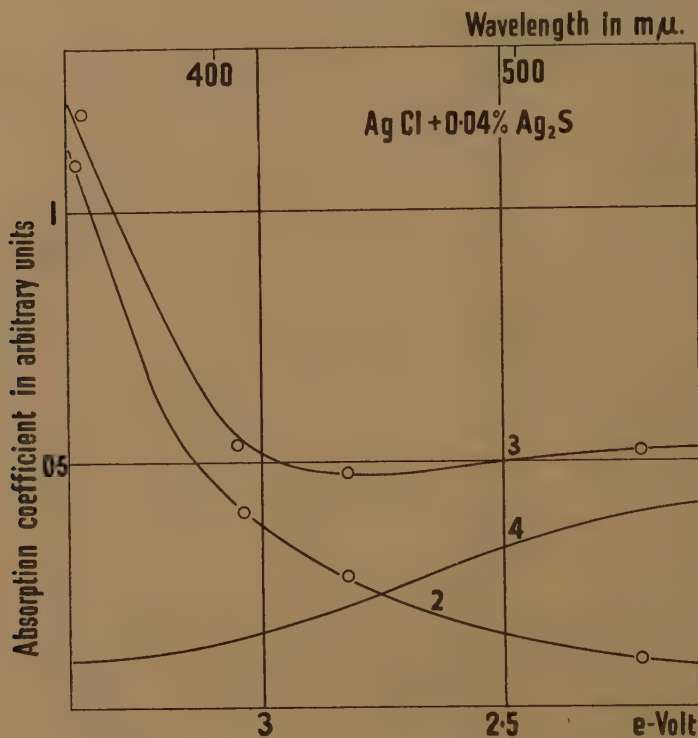


The effect of annealing and slowly cooling the specimens. Curve 2. The absorption spectrum of the quenched specimen. Curve 5. The absorption spectrum of the annealed specimen. Curve 6. The absorption spectrum of a thin silver sulphide layer.

When the quenched specimens were irradiated, either in the absorption band of the pure lattice or in the long wave tail of the impurity band, new absorption bands appeared in the same positions as those which were found by Hilsch and Pohl (1930) in irradiated pure silver halide crystals, and assumed by them to be due to the presence of separated colloidal silver. The results are included in fig. 1. When the mixed crystal specimens are irradiated in the long wave tail of the impurity band there is, according to Stasiw and Teltow, no measurable change in the absorption coefficient in that region. The experimental evidence for the conclusion is shown in fig. 3.

The discussion of the absorption band due to colloidal silver was based on Savostianova's papers (1934, 1935), in which the approach is not entirely satisfactory, as pointed out by Berg (1938). The absorption band with a maximum near 6200 Å which appears during photolysis is assumed by Stasiw and Teltow to be due to the formation of colloidal silver particles of mean diameter about 70 m μ , the particles having a range of sizes and containing on the average 10^7 silver atoms *. Provided

Fig. 3.



The absence of any decrease in the absorption coefficient in the impurity band following irradiation.

that the size is assumed to be not greater than this, so that the specks are "transparent", it is possible to deduce the quantity of separated silver from the characteristics of the absorption band, and from this to find the quantum efficiency. Values of 0.5 for the quenched mixed crystals with silver chloride and of 0.2 for the quenched silver bromide mixed crystals were obtained by Stasiw and Teltow compared with their much lower values for the pure silver halides of between 0.05 and 0.005. Low quantum efficiencies were also found for mixed crystal specimens which had been annealed and then slowly cooled.

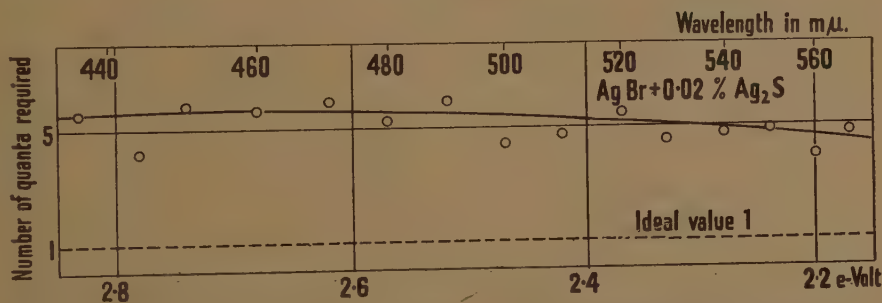
* In our view, the particles which are mainly responsible for this absorption band consist of aggregates of F centres whose maximum dimension is about 20 Å.

In silver bromide crystals in the spectral region between 4300 Å and 5700 Å the quantum yield is almost independent of the wavelength, showing that absorption in the lattice and in the impurity band contributed equally in the formation of colloidal silver. The result is shown in fig. 4. The maximum amount of silver which could be liberated was about equal to the number of silver sulphide molecules in the mixed crystal.

The photolytic separation of colloidal silver occurred uniformly throughout the bulk of the quenched mixed crystals, contrasting sharply with the behaviour of pure silver halide crystals, in which surface separation is mainly observed (Haynes and Shockley 1948).

The deep green to black crystals containing separated colloidal silver were restored to their original clear condition by reheating and quenching. This result shows that although a photochemical process has occurred with a high quantum efficiency, there has been no escape of halogen from the crystal.

Fig. 4.



The dependence of the number of quanta required to effect the separation of one silver atom on the wavelength of the incident illumination.

When a clear quenched silver bromide crystal was exposed to bromine vapour at room temperature, the brown coloration due to the impurity absorption band was bleached. A sharply defined interface moved into the crystal from the surface, leaving behind a clear yellow zone. At intermediate stages this edge zone was insensitive to light, like pure silver bromide, while the unchanged middle zone retained its high sensitivity.

The deep green colour of the photochemically darkened silver bromide crystals was bleached at approximately the same rate as the brown colour of the clear crystals, by exposure to bromine vapour. When a darkened crystal was removed from the bromine with a central zone unbleached, this zone could be bleached as above by warming the crystal to 250°C.

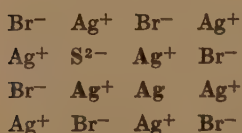
Similar impurity bands, extending the absorption towards the red, appeared when mixed crystals were made from silver halide melts to which small quantities of silver oxide, silver nitrate, silver selenide, silver cyanide or silver cyanate had been added.

The ionic conductivity of the mixed crystals was measured in the temperature region between 300 and 400° C. and, as found previously by Koch and Wagner (1936), no measurable increase in the conductivity due to the addition of silver sulphide was observed. From this Koch and Wagner had concluded (we believe incorrectly) that mixed crystal formation did not occur. The photochemical evidence of Stasiw and Teltow shows that, in crystals which have been quenched from the temperatures at which the ionic conductivity was measured, there is an impurity absorption band due to the incorporation of silver sulphide in the lattice.

3. INTERPRETATION OF THE EXPERIMENTAL RESULTS SUGGESTED BY STASIW AND TELTOW.

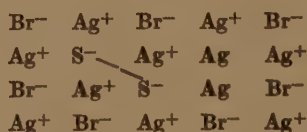
In discussing the interpretation of their results, Stasiw and Teltow placed considerable emphasis on the observation, referred to in the previous section and shown in fig. 3, that the absorption coefficient in the long wave tail of the impurity band did not decrease during irradiation even though photolytic silver was liberated. They therefore attempted

Fig. 5.



Model of an $\text{Ag}(\text{Ag}_2\text{S})$ -complex
proposed by Stasiw and Teltow.

Fig. 6.



Alternative model of an
 $\text{Ag}_2(\text{Ag}_2\text{S}_2)$ complex.

to devise models where the impurity centre would be regenerated after the absorption process. In these, a light sensitive complex was assumed to be formed by the close association of silver atoms with built-in silver sulphide molecules; two groupings of silver atoms with silver sulphide molecules were suggested. In the first, fig. 5, a silver atom occupied the site of a halide ion. This satisfied the requirement that the crystal as a whole should be electrically neutral, but necessitated an excess of silver over the stoichiometric composition. It was assumed that this excess could easily be introduced during the preparation of the crystals. In the second arrangement, fig. 6, two silver atoms were grouped together with two silver sulphide molecules, this avoiding the necessity for any excess of silver over the stoichiometric composition.

On the basis of these models of the impurity centre they proposed, the following interpretation of their results:

Absorption in the impurity band ejects an electron from a doubly charged sulphide ion. The electron diffuses through the lattice until, together with an interstitial silver ion, it is captured by a colloidal silver particle forming a silver atom. An electron is now transferred from the

silver atom in the silver sulphide adsorption complex to the singly charged sulphide ion, reforming a doubly charged sulphide ion.

With this mechanism, an electron is ejected from the complex by the absorption of a light quantum, and as the regeneration of the complex involves a silver atom which did not contribute to the absorption in the long wave region, it is assumed that the absorption coefficient in that region will not change.

When the illumination had a wavelength lying in the absorption region characteristic of the silver halide lattice, they proposed the following scheme for the photochemical process :

A bromide ion absorbs a light quantum giving an electron and a bromine atom or positive hole in the full band of the crystal. The electron diffuses through the lattice until, together with an interstitial silver ion, it is captured by a colloidal silver particle forming a silver atom. The bromine atom or positive hole diffuses to one of the light-sensitive silver sulphide complexes, where it reacts with the adsorbed silver atom, giving a bromide ion and a silver ion.

The absorption characteristics of the complex are assumed to remain unchanged in the long wave region. They also assumed that, in the absence of colloidal silver particles, the adsorption complexes could themselves act as nuclei for silver separation.

4. CRITICISM OF THE MODEL OF THE IMPURITY CENTRE SUGGESTED BY STASIW AND TELTOW.

Bodenstein (1941) has criticized this interpretation on a number of chemical and thermodynamic grounds. There are further objections. It is assumed that the absorption of a light quantum ejects an electron from an S^{--} ion, leaving an S^- ion. If the silver atom is in the position shown in fig. 5, the transfer of an electron from it to the S^- ion is unlikely to occur. To a crude approximation, the electrostatic energies for an electron at the sites occupied by the sulphide ion and by the silver atom will be equal. The removal of the electron from the silver atom will require an amount of work approximating to the ionization energy of the atom; moreover, the electron affinity of a sulphide ion for a second electron is negative. The binding energy of the electron in the silver atom is thus very much greater than that of the electron in the S^{--} ion, and it is because of this that the silver atom does not contribute to the absorption in the long wave region.

It does not seem possible to account for the internal separation of "colloidal" silver with either of these models. Further, the model does not explain why the absorption band due to the impurity centres, which is apparently completely bleached when the quenched mixed crystals are exposed to bromine vapour at room temperature, is not bleached when the bromine atoms formed by the absorption of light quanta react with the centres. At room temperature, sulphide ions are unlikely to diffuse from the lattice, and if they are responsible for the impurity absorption band it should not be bleached by exposure to bromine vapour.

5. AN ALTERNATIVE MODEL FOR THE ABSORPTION CENTRE IN MIXED SILVER HALIDE CRYSTALS.

In the past, it has been assumed that the ionic conductivity of pure silver halide crystals was due to the presence of Frenkel defects, which are silver ions in interstitial positions with a corresponding number of vacant silver ion lattice sites (Wagner and Beyer 1936, Just 1938). A review of all the evidence suggests that this conclusion has not been fully established and, indeed, much indirect evidence, which will be discussed in detail in a later paper, favours the conclusion that the defects in silver chloride and silver bromide are of the Schottky type, consisting of vacant cation lattice sites with an equal number of vacant anion lattice sites. Frenkel defects may possibly occur in the β and γ forms of silver iodide, and X-ray structure determinations have shown that the silver ions occupy interstitial positions in α silver iodide (Strock 1934).

Anastasevich and Frenkel (1941) have discussed the photolysis of silver bromide crystals assuming the presence of Schottky defects, and pointed out that the photoelectrons are likely to be captured by vacant anion lattice sites, thus forming F centres. They do not, however, provide any evidence for the existence of F centres. They suggest that, under conditions (to be discussed) where F centres are mobile, they can flock together to form colloidal silver particles. This we believe to be correct. It is an alternative mechanism to that of Gurney and Mott and will provide for the formation of *internal* latent image. Both may occur under appropriate conditions, the Gurney-Mott mechanism when electrons are trapped by colloidal metal at the surface, and the aggregation of F centres providing nuclei for surface or internal silver separation in crystals without many electron traps.

One piece of evidence for the existence of Schottky defects may be mentioned here. Mixed crystals of silver chloride or silver bromide with cadmium halides are *relatively* insensitive to light (Burgers and Kooy 1948). Assuming Schottky defects, such crystals are without vacant anion sites, and thus will not form F centres on illumination. Apart from the action of accidental surface traps, there is thus no possibility of photolysis.

As regards the mobility of the vacant sites, Tubandt's (1932) observation from measurements of ionic conductivity that the cation transport number is unity forces us to assume a low mobility for the vacant anion sites at room temperature. This is also known to be the case with sodium chloride at low temperatures*.

* Neutral pairs formed by the association of a vacant anion site with a vacant cation site should have a higher mobility than the vacant anion site alone. This question has been considered by Dienes (1948). Under equilibrium conditions, the concentration of such pairs in silver halide crystals is likely to be very small (Mott and Gurney 1940, p. 42), but the diffusion of pairs may play a part for reactions depending on ionic migration processes in quenched crystals where there is a non-equilibrium concentration of lattice defects.

Turning now to the mechanism by which silver sulphide is incorporated in the lattice, there seem to be two possibilities :

- (a) Through the incorporation of an S^{--} ion and an interstitial silver ion.
- (b) Through the incorporation of an S^- ion, replacing a halide ion, and an F centre.

In case (a), if the two were separated from each other, as is likely at elevated temperatures, an enhanced conductivity would result, which was not observed by either Koch and Wagner or Stasiw and Teltow. An S^- ion together with an interstitial silver atom is not possible either, since the atom would be dissociated with a very low activation energy, giving electronic conduction. This is shown by calculations made in this laboratory by J. H. Simpson (1949). Further evidence against the association of a silver atom or ion in an interstitial position with a sulphide ion is provided by the observation of Stasiw and Teltow of the internal separation of photolytic silver in these crystals ; in either case there is no reason for supposing that electrons in the conduction band would be effectively trapped anywhere but at the surface of the crystal, as in the pure halides.

The model which we propose, case (b) above, seems to be undoubtedly the correct one. On the basis of an empirical extrapolation, the electron affinity of a sulphur atom for one electron is $+2.06$ e.v. (Massey 1938), while the electron affinity of sulphur for two electrons deduced by Mayer and Maltbie (1932) from the Born-Haber cycle is $-(3.9 \pm 2.2)$ e.v. This means that sulphur has a negative affinity for a second electron of about 6 ± 2 e.v., compared with the positive affinity of about 3 e.v. for halogens. It is clear, then, that much energy will be gained if the electron is transferred from S^{--} to a vacant anion lattice site forming an F centre and leaving a singly charged sulphide ion occupying a halide ion lattice site. Although a sulphide ion with an effective charge close to two electronic units can probably exist in the environment of positive divalent ions provided by a barium sulphide or lead sulphide crystal, there seems to be no reasons for supposing that this will be possible in a silver halide crystal. The binding energy of an electron to a vacant anion site in a silver chloride or bromide crystal is not yet known experimentally, but by comparison with the properties of the F centres in lithium halide crystals it seems likely that it will lie between 2.5 and 3.5 electron volts. The formation of F centres in mixed crystals of silver halides with silver sulphide is thus consistent with energetic considerations. There does not seem to be any reason for assuming the formation of association complexes between F centres and singly charged sulphide ions ; they will probably behave as independent units in the lattice. As regards the F centres, which will not lose their electrons at room temperature, we assume them to be **immobile** except when a vacant anion site moves to an adjacent or near-by site, so that the electron can jump to it by tunnel effect.

We therefore suggest that the properties of the quenched mixed crystals of silver chloride and silver bromide with silver sulphide, which have been outlined in Section 2, are due to the presence of F centres. Using this model, we will discuss the experimental results of Stasiw and Teltow in the next section in the same order as in Section 2.

6. AN ALTERNATIVE INTERPRETATION OF THE RESULTS OF STASIW AND TELTOW.

The absorption bands of fig. 1 are assumed to be due to the presence of F centres. Clear mixed crystals which show these bands can be produced only by rapid quenching from temperatures near the melting point. This behaviour has also been observed by Rögner (1937) with alkali halide crystals containing excess alkali metal. If the crystals are slowly cooled, the F centre band is accompanied by a band which has been assumed to be due to the presence of colloidal alkali metal in the crystal. We now believe that this band is at least partly due to the presence of F centre aggregates. This will be discussed in Section 7.

The colour of the quenched mixed crystals may therefore be attributed to absorption of light in the long wave tail of an F centre band whose maximum is in the region of absorption characteristic of the lattice. The absorption coefficient in the long wave region should be temperature dependent (Mollwo 1933). An absorption band due to the presence of monovalent sulphide ions is also likely to contribute to the absorption coefficient in the long wave region.

When the mixed crystals are annealed and then very slowly cooled, silver sulphide may separate as a distinct phase. However, since rapid quenching results in the formation of F centres, somewhat slower rates of cooling may produce F centre aggregates. The initial stage will be the formation of a double centre, in which two electrons are associated with two adjacent vacant anion lattice sites. In the double centre the electrons are able to move in a larger volume than in the isolated centres, and they will therefore have lower energies in their ground states. By combination with further F centres larger aggregates may be formed which may be linear, two-dimensional or three-dimensional. In such aggregates the configuration of the silver ions will approximate to that in the silver halide lattice. In the larger aggregates, rearrangement with the formation of colloidal silver particles should be possible. The long wave extension of the absorption band in the annealed crystals, shown in fig. 2, is therefore probably due partly to the presence of silver sulphide in the crystal as a distinct phase, and partly to F centre aggregates and colloidal silver particles. Further experimental work is required before this question can be finally settled.

Illumination with wavelengths corresponding to the long wave tail of the absorption band will result in the ejection of an electron into the conduction band of the crystal from either a sulphide ion or an F centre. It seems that before an electron in the full band of the crystal could be

transferred to the sulphur atom remaining after the former absorption process, it would have to acquire an energy increment corresponding to the difference between the electron affinities of the halogen and sulphur. As this increment is considerably greater than kT , the sulphur atom may remain in the lattice as an electron trap which might eventually capture an electron from the conduction band, thus reducing the efficiency of the photochemical process. The ejection of an electron from an F centre into the conduction band will leave a vacant anion lattice site. Irradiation with light of wavelengths in the lattice absorption band will raise electrons to the conduction band and leave mobile positive holes in the full band.

The interaction of electrons and positive holes with trapping sites will now be considered. Electrons may be captured by vacant anion sites forming F centres. The F centres are themselves potentially capable of capturing a second electron to form F' centres (Pohl 1937), but such centres will probably be dissociated at room temperature. Relatively deep traps for electrons will be provided by F centre aggregates (and colloidal silver particles derived from them), and these we believe to provide the initial traps of the photolytic process—*i. e.* the sensitivity specks of the silver halide grains in photographic emulsions. When surface colloidal silver particles trap electrons the charge may be neutralized by the migration of silver ions, the original mechanism of Gurney and Mott only probably taking place through the motion of vacant cation sites. When, however, aggregates which do not lie entirely on the surface capture electrons, the only possibility for neutralizing their charge is the motion towards them of the much less mobile vacant anion lattice sites. These questions will be discussed in greater detail in the next section.

The absorption band due to the reaction product, whether F centre aggregates or colloidal silver, is superimposed on the long wave tail of the F centre absorption band. The band due to the reaction product can be bleached by light (Hilsch and Pohl 1930), through the ejection of electrons into the conduction band of the crystal. If these are captured by vacant anion sites, F centres will be formed. The photochemical reaction is therefore reversible and, under these circumstances, it will not be possible to destroy the F centre band completely through illumination. Further, one or two dimensional aggregates (rows or plates) will be dichroic, absorption with the electric vector perpendicular to the length or to the surface corresponding to an absorption band similar to the F centre absorption band, and with the electric vector parallel to the length or to the surface to an absorption band with a maximum at longer wavelengths. We believe that these considerations, together with the contribution from the monovalent sulphide ions, account for the observation that the absorption coefficient in the long wave tail was apparently not appreciably diminished during irradiation.

The quantum efficiency will be determined by the relative probabilities for the capture of electrons by vacant anion lattice sites and by F centre

aggregates, as well as by the efficiency with which positive holes are captured by isolated F centres. In the absence of F centres, as in the pure silver halides or in the annealed mixed crystals, positive holes probably recombine with the trapped electrons, accounting for the low quantum efficiencies.

The quantum yield should be almost independent of wavelength, as shown in fig. 4, because the processes following the absorption of a light quantum are essentially the same whether absorption is by an F centre or by a bromide ion of the lattice. If the electron is ejected from an F centre, it will move through the lattice until it is captured either by a vacant anion site or by an F centre aggregate. A vacant anion site will remain. An electron ejected from a bromide ion will behave in the same way and the mobile positive hole will probably be captured by an F centre forming a vacant anion lattice site. These additional vacant anion lattice sites, which will be in excess of the equilibrium concentration, will disappear through secondary processes. If one F centre is formed for each molecule of silver sulphide added, the number of silver atoms which can be aggregated by light with a high quantum efficiency should be equal to the number of silver sulphide molecules.

If the photochemical process involves the aggregation of F centres, photolytic silver can separate internally without creating severe lattice strains. In pure silver halides, surface separation alone is observed if well-annealed single crystals are used for the experiments. In these crystals, vacant anion lattice sites may initially trap electrons forming F centres which later react with positive holes so that there is no resultant internal photochemical process. At the surface of the crystal, when electrons are trapped by vacant anion lattice sites, there will be a certain probability for the escape of bromine. One F centre can be formed for each bromine atom which escapes from the crystal. By reheating, it will not be possible to restore to its original clear condition a pure silver halide crystal which has been deeply coloured because the separation of "colloidal" silver involves a loss of bromine. On the other hand, the photochemical process in the quenched mixed crystals does not result in any change in the total chemical composition of the crystals.

When a quenched mixed crystal is exposed to bromine vapour, the long wave tail of the absorption band is bleached, but few quantitative details are given by Stasiw and Teltow. It may be assumed that when bromine is adsorbed on the surface of the crystal, the molecules are dissociated and partially transformed into bromide ions. The positive holes which are simultaneously formed migrate into the crystal and combine with F centres, forming vacant anion lattice sites. This must create in the reaction zone a positive space charge equivalent to the negative surface charge associated with the additional surface bromide ions. These charges will then be neutralized by ionic migration, the details depending on the relative mobilities of vacant cation and vacant anion lattice sites. Vacant cation sites will be formed at the surface where

the bromine is adsorbed and migrate into the crystal. They will combine with the vacant anion lattice sites derived from the F centres, forming associated neutral pairs of vacant sites. Since these will be in excess of the equilibrium concentration of such pairs, and have a high mobility, they will migrate to the surface and disappear. The migration of the vacant anion sites to the surface, if the mobility were big enough, would produce an equivalent result. In either case the rate of reaction will be determined by the ionic diffusion rate. When the F centres have been destroyed by these processes, the subsequent photochemical reactions in the crystals will have the low quantum efficiency associated with the pure silver halides. When the crystals are exposed to bromine vapour, the F centre band should be completely bleached, and this is in no way inconsistent with the behaviour on irradiation. The bleaching of the photochemically darkened crystals by exposure to bromine is due essentially to the same processes. Before illumination, the quenched crystals behave as though they contained excess silver in the form of F centres, and after illumination excess silver in the form of F centre aggregates. The treatment with bromine removes this silver to form silver bromide, and after treatment the crystal must contain an excess of bromine compared with its original composition, there being one excess bromide ion for each singly charged sulphide ion. At room temperature the sulphide ions will remain in fixed positions in the structure and will still make the same contribution to the absorption spectrum. The experiments thus seem to prove that the optical properties of the crystals are not determined by the sulphide ions. The fact that the rate of bleaching is approximately the same before and after photochemical darkening suggests that the state in which the silver exists in the two cases is not very different. If the photochemical process resulted in the formation of large internal colloidal silver aggregates, the rate of bleaching and the sharpness of the boundary between attacked and unattacked parts of the crystal would be different for clear and darkened crystals.

Mixed crystals of silver halides with silver oxide should also show absorption bands due to the presence of F centres. The same reasoning will apply to the behaviour of the oxide ion in the silver halide lattice as has been applied to the sulphide ion.

The estimated position of the maximum of the F centre absorption band suggests that the F centres may not be thermally dissociated below the melting point of the silver halide. If this were the case, as they are neutral centres, their presence in the crystal would not result in any increase in the electrical conductivity.

7. THE PROPERTIES OF THE F CENTRES AND F CENTRE AGGREGATES IN SILVER HALIDES.

In this section we shall first summarize the properties of isolated F centres in silver halides and then consider the manner in which the F centres combine to form aggregates and the properties of the aggregates.

We believe that these properties provide the key to the explanation of the differences between the sensitivity speck, the unstable latent sub-image, the stable latent sub-image and the latent image speck in the sensitive silver halide grains of photographic emulsions, and we outline a possible new mechanism of latent image formation.

The F centre consists of an electron trapped in the field of a vacant anion lattice site (De Boer 1937, Mott 1937). Very recent theoretical calculations by J. H. Simpson (1949) in this laboratory have given a value of 2.1 electron volts for the binding energy of an electron in an F centre in silver bromide. The experimental evidence suggests that the binding energy is greater than this, probably lying between 2.5 and 3.5 electron volts. The maximum of the F centre absorption band in silver bromide should therefore occur at a wavelength between 3500 and 5000 Å, but particular interest centres on the long wave tail, extending to at least 6200 Å. The absorption coefficient in this long wave tail should be temperature dependent. We suggest that the extension of the absorption region to longer wavelengths which has been observed by Eggert and Kleinschrod (1940) in silver bromide photographic emulsions is partly due to the presence of F centres introduced during the after-ripening process.

The absorption of a light quantum by an F centre will cause the electron to jump from an s state to a p state, from which it can be raised to the conduction band if, within the life-time of the excited state, it can acquire the necessary additional energy from lattice vibrations (Mott and Gurney 1940, pp. 113 and 133). We assume that this is possible at room temperature, the absorption of light quanta by F centres freeing electrons and leaving vacant anion lattice sites. The quantum efficiency should fall at very low temperatures (Glaser and Lehfeldt 1936), but no measurements of photoconductivity on the quenched mixed crystals which would establish these properties have yet been recorded.

The results of the measurements of electrical conductivity show that the F centres in silver halides are not thermally dissociated into an electron and a vacant anion lattice site at temperatures below their melting points. Except at very low temperatures, isolated F centres are unlikely to combine with a second electron from the conduction band to form F' centres on account of the low binding energy for the second electron. They can, however, combine with positive holes in the full band to form vacant anion lattice sites.

The migration of F centres by thermal diffusion processes will involve anion displacements. Anastasevich and Frenkel (1941) suggested that migration could occur at room temperature and took place through successive displacement of bromide ions adjacent to the centre. If this were the case, the quenched silver halide silver sulphide mixed crystals studied by Stasiw and Teltow would not be stable, since the F centres would form aggregates at room temperature. Moreover, the results of experimental investigations of the photolysis of silver halide precipitates and the properties of the silver halide grains in high-speed emulsions

can only be interpreted if the mechanism by which the F centres move depends on the presence of additional vacant anion lattice sites; these results cannot be understood if the mechanism put forward by Anastasevich and Frenkel is correct. This argument will be presented in a later paper. We will meanwhile assume that the F centre remains at rest until a slowly moving vacant anion site comes up to within a few lattice sites of it. The electron may then jump across (by tunnel effect) and the original vacant site diffuse away. The evidence indicates that the thermal diffusion constant for F centres is very small at room temperature, corresponding to the low transport number of the vacant anion site.

F centres diffusing thermally by this mechanism will combine to form aggregates. The smallest aggregate will consist of two F centres in which the vacant anion sites may be separated by a distance a or by a distance $\sqrt{2}a/2$, where a is the side of the unit cell. A third F centre combining with this double aggregate may extend it in a $\langle 100 \rangle$ or a $\langle 110 \rangle$ direction or may form a two-dimensional aggregate. These triple aggregates will provide nuclei for further growth in one, two or three dimensions. We believe that the diffuse bands and diffuse spots in the Laue X-ray diffraction photographs of silver chloride obtained by Burgers and Tan Koen Hiok (1946) provide evidence for the presence of groups of linear F centre aggregates along $\langle 100 \rangle$ directions and of two-dimensional F centre aggregates in $\{100\}$ planes. The diffuse diffraction patterns which do not completely disappear at low temperatures would then owe their origin to the presence in the crystal of vacant anion lattice sites regularly arranged in rows and sheets, the trapped electrons not appreciably altering the scattering. This model appears to be consistent with all their published observations. The results show that the linear aggregates are about $4a$ units in length. They may have been originally present in the crystal or have been formed during its exposure to the X-ray beam.

We attribute the latent image band in silver chloride, observed by Hilsch and Pohl (1930) and by Löhle (1933), to the formation in the crystal of small one- and two-dimensional aggregates of F centres of this type and of approximately these dimensions. Such aggregates will also be formed during photolysis in the $\text{AgCl-Ag}_2\text{S}$ mixed crystals which show the same absorption band. In these aggregates, the configuration of the silver nuclei will be the same as in the silver chloride lattice and not that of metallic silver. Berg (1943) has already presented evidence in favour of the conclusion that this absorption band is due to small colloidal silver particles; we propose a definite model for the particles which will have to be confirmed by further experimental work. The latent image band in large single silver bromide crystals and in silver bromide silver sulphide mixed crystals will be due to the presence of similar aggregates of F centres.

The light absorption characteristics of the aggregates will depend on their size and on the orientation of the electric vector of the incident beam. For a given size there will be two distinct absorption maxima, one when the electric vector is perpendicular and the other when it is parallel to the axis or surface of the aggregate. The position of the first maximum

should vary only slowly with aggregate size and should always lie on the short wave side of the second; this second maximum will be displaced steadily towards longer wavelengths with increasing size. These properties can be understood in a general way by assuming that the light quanta are absorbed by electrons confined within a "potential box" of which the potential, on account of the high dielectric constant, does not rise abruptly at the boundaries. The variation of potential along a transverse direction will to a first approximation be independent of the size of the aggregate. The energy of the light quantum, with electric vector parallel to the axis or to the surface, required to raise an electron from the highest occupied energy state to the first unoccupied state to which an optical transition is permitted will decrease as the size of the aggregate increases. The superposition of the absorption due to groupings of F centres and of linear and two-dimensional aggregates with a range of sizes oriented along all the $\langle 100 \rangle$ directions and lying in all the $\{100\}$ planes will account for the observed characteristics of the latent image band.

The results of the investigations with mixed crystals of silver halides and silver sulphide show that the uncharged F centre aggregates are thermally stable at room temperature but are dispersed when crystals containing them are heated to several hundred degrees. It seems that heating pure silver halide crystals containing separated photolytic silver results in the formation of metallic colloidal silver particles, which are not re-dispersed by further prolonged heating, because rapid quenching of the crystal apparently does not produce F centres.

The absorption of a light quantum by an F centre aggregate (in AgCl red light when the electric vector is parallel to the axis or surface of the aggregate, for example) will excite an electron to a higher energy state from which it may be raised to the conduction band by acquiring the necessary additional dissociation energy from lattice vibrational energy. At a sufficiently high temperature, the remaining *small* positively charged F centre aggregate will dissociate into a neutral aggregate and a vacant anion lattice site. The absorption coefficient for light of polarization and wavelength identical with that of the incident beam, which was responsible for the excitation of the aggregate, will thus be diminished; and simultaneously the absorption coefficient for light of shorter wavelength with the same polarization direction will be increased because, parallel to that direction, the aggregates will decrease in size and the liberated electrons will finally all be trapped by isolated vacant anion lattice sites or by aggregates oriented in one of the other two possible directions. These properties provide, we believe, the explanation of the bleaching of the latent image band by red light (Hilsch and Pohl 1930) and of the Weigert effect (Cameron and Taylor 1934, Freundlich 1936). In connection with the latter effect, we would emphasize that F centre aggregates of the type which we envisage are dichroic by nature and have light absorption characteristics formally similar to those of linear and planar polyatomic organic molecules.

During the illumination of a crystal containing them, F centre aggregates may capture electrons from the conduction band. We consider that a negative charge on an internal F centre aggregate can only be neutralized by a slowly moving vacant anion lattice site ; also at room temperature, particularly if the concentration of vacant anion lattice sites is small, the electron may be ejected thermally from a *small* aggregate before a vacant anion site moves into its neighbourhood. If this happens, the electron is likely eventually to be captured by a vacant anion lattice site. Even when the electron has not already been lost through thermal excitation, it will be transferred to an approaching vacant anion lattice site by tunnel effect when the site is still several inter-ionic distances away. Direct combination would add one F centre to the aggregate, but it seems that this cannot occur with a small aggregate. Instead, the primary result of the absorption of light is a redistribution of the isolated F centres, a number being transferred to the immediate neighbourhood of the aggregates. Such groupings will slowly disperse through thermal diffusion, some of the F centres being added to the aggregates and the remainder moving away. As the dimensions of the aggregates increase, due to these photolytic and thermal diffusion processes, the thermal activation energy, W , required to free the trapped electron will increase, and corresponding with this, at a given temperature, the life-time of the negatively charged aggregate, $t=t_0 \exp(W/kT)$, will increase. These changes will increase the efficiency with which the negative charge can be neutralized by a vacant anion lattice site on or in the immediate neighbourhood of the aggregate.

We assume that when a light quantum (infra-red) is absorbed by a negatively charged aggregate, the electron is ejected directly into the conduction band, as occurs for an F' centre. Negatively charged aggregates may therefore be dissociated at all temperatures into an electron and a neutral aggregate, but the neutral aggregates themselves, as mentioned above, are not dissociated optically unless the energy of the electron acquires a further increment from lattice vibrational energy within the life-time of the excited state.

An aggregate of n F centres can combine with a vacant anion lattice site to form a positively charged aggregate of $(n+1)$ F centres. The thermal activation energy for its dissociation into a neutral aggregate and a vacant anion lattice site, $V_{(n+1)}$, will increase with $(n+1)$. Below a certain critical size the life-time, $t=t_0 \exp(V_{(n+1)}/kT)$, of the positively charged aggregate will not be significant ; above this size, the life-time will increase rapidly with $(n+1)$. The critical size, which will depend on the temperature and on the vacant anion lattice-site concentration, will be reached by the further addition of only one F centre or vacant anion lattice site to an aggregate. At a given temperature it will be smaller the higher the vacant anion lattice site concentration, which in photographic emulsions depends on the silver and hydrogen ion concentrations of the medium surrounding the silver halide grains (Mitchell 1948).

The positively charged aggregate will provide a relatively deep electron trap and is less likely to react with a positive hole than a neutral aggregate.

The small linear and two-dimensional aggregates, and groupings of aggregates and F centres, whose properties we have been considering cannot actually lie on the surface of the silver halide crystals, although they may lie very near the surface. They are therefore always essentially *internal* aggregates and groupings and must increase in size through the migration of vacant anion sites.

The growing positively charged aggregates will eventually reach a second critical size at which they break away from the silver halide matrix and collapse to form colloidal silver particles with the lattice constant of metallic silver. This will certainly happen on prolonged photolysis and during the later stages of chemical development. The colloidal silver particles which are formed near the surface in this way will probably be exposed during the rearrangement; they will act as efficient electron traps, and their negative charge can now be neutralized by the movement of silver ions from near-by lattice sites, vacant cation sites migrating away. This is the original Gurney-Mott mechanism, adapted to a model with Schottky defects instead of Frenkel defects. Energetic considerations suggest that the colloidal silver particles will be positively charged when they are in thermal equilibrium with the silver halide. After the neutralization of the positive charge by an electron from the conduction band, they can become positively charged again through the movement of a silver ion by the above mechanism.

These properties lead to the following tentative model for the sensitivity speck, the unstable latent sub-image, the stable latent sub-image and the latent image speck. The normal sensitivity speck, produced through the chemical after-ripening of a photographic emulsion, is a stable aggregate of a few F centres. The unstable latent sub-image (Berg 1947) is the grouping of F centres round the sensitivity speck, which results from exposure under certain conditions. In the dark, the speck may increase slowly to the critical size through the thermal diffusion of the group of F centres, both the critical size and the diffusion rate being determined, at a given temperature, by the local vacant anion lattice-site concentration. The small sensitivity speck may also grow directly during exposure by capturing an electron from the conduction band while it is positively charged through transitory association with a vacant anion lattice site. Above the first critical size, the aggregate is always positively charged, and forms the stable latent sub-image. We estimate that, under the conditions prevailing in photographic emulsions, the aggregate of critical size will have less than ten missing halide ions. The photolytic process changes essentially at this point. The positively charged aggregate now captures an electron from the conduction band to form a stable neutral aggregate which will then combine with a further vacant anion lattice site and become positively charged again. The ionic migration step thus occurs partly before the exposure instead of entirely during or

after it. More than one vacant anion lattice site may be associated with large latent image specks. The positively charged speck will be surrounded by a negative space charge due to a statistical excess of vacant cation lattice sites over vacant anion lattice sites in its immediate neighbourhood, the system behaving as a small condenser.

The sensitivity speck may be built up to just below the lower critical size by a carefully controlled uniform hypersensitizing pre-exposure of the emulsion. This will produce unstable latent sub-image specks consisting of a sensitivity speck with a number of isolated F centres in its immediate neighbourhood. If the emulsion is now exposed to the optical image formed by a lens system, the absorption of very few quanta will cause the speck to change from a neutral speck to a positively charged speck, and beyond this critical size all the F centres of the latent sub-image are likely to be transferred to the speck by further exposure. In a similar way, an unstable latent sub-image produced by exposure to an optical image may be converted into a stable latent image by a carefully controlled uniform post-exposure, a process referred to as latensification (Berg 1947, p. 277), before the sensitivity specks of any of the unexposed grains reach the critical size.

We consider that the silver halide grain is potentially developable when, at the temperature of the developer and with the vacant anion lattice-site concentration determined by its hydrogen ion and silver ion concentrations, the latent image speck exceeds the first critical size. Under these conditions, electrons may be transferred by tunnel effect from adsorbed negative developer ions to the positively charged latent image specks near the surface of the grain, but not to the smaller uncharged sensitivity specks. In the developer, the latent image speck increases slowly in size by the same mechanism as in the photolytic process, namely the migration of vacant anion lattice sites. The initial rate of development will depend on the size and proximity to the surface of the latent image speck—on the oxidation-reduction potential of the developer and on the concentration and mobility of the vacant anion lattice sites on the surface layers. Development will continue in this way until a two-dimensional speck at the surface reaches the second critical dimension at which it breaks away from the silver halide matrix and forms an exposed colloidal silver particle. Beyond this stage, rapid three-dimensional growth will occur by the Gurney-Mott mechanism, a thread or ribbon of silver being pushed out from the surface.

This model is strongly supported by much available photographic evidence. Many details will have to be confirmed by calculations and experimental work, which will require a considerable time.

ACKNOWLEDGMENTS.

The author wishes to thank Professor N. F. Mott for very frequent discussions during the development of the ideas contained in this paper, which depends a great deal upon his earlier work with Dr. R. W. Gurney.

He is also grateful to Dr. W. F. Berg for valuable assistance in working out the application of the conclusions to a number of aspects of the photographic process.

REFERENCES.

- ANASTASEVICH, V. S., and FRENKEL, J., 1941, *J. Exper. and Theoret. Phys. U.S.S.R.*, **11**, 127.
- BERG, W. F., 1938, *Trans. Faraday Soc.*, **34**, 889; 1943, *Ibid.*, **39**, 115; 1946-7, *Reports on Progress in Physics*, **11**, 248.
- BODENSTEIN, M. L., 1941, *Abhand. der Preuss. Akad. der Wiss. Math. Phys. Kl.*, No. 19.
- DE BOER, J. H., 1937, *Rec. Trav. Chim. Pays-Bas*, **56**, 301.
- BURGERS, W. G., and KOOY, J. N., 1948, *Rec. Trav. Chim. Pays-Bas*, **67**, 16.
- BURGERS, W. G., and TAN KOEN HIOK, 1946, *Physica*, **11**, 353.
- CAMERON, A. E., and TAYLOR, A. H., 1934, *J. Opt. Soc. Amer.*, **24**, 316.
- DIENES, G. J., 1948, *J. Chem. Phys.*, **16**, 620.
- EGGERT, J., and KLEINSCHROD, F. J., 1940, *Z. wiss. Photogr.*, **39**, 155.
- FREUNDLICH, H., 1936, *Photo. J.*, **76**, 49.
- GLASER, G., and LEHFELDT, W., 1936, *Nachr. Ges. Wiss. Göttingen, Math. Phys. Kl.*, **2**, 91.
- GURNEY, R. W., and MOTT, N. F., 1938, *Proc. Roy. Soc. A*, **164**, 151.
- HAYNES, J. R., and SCHOCKLEY, W., 1948, *Report on Conference on Strength of Solids* (London: The Physical Society), p. 151.
- HILSCH, R., and POHL, R. W., 1930, *Z. Physik*, **64**, 606.
- JOST, W., 1938, *Trans. Faraday Soc.*, **34**, 860.
- KOCH, E., and WAGNER, C., 1936, *Z. Phys. Chem. B*, **38**, 295.
- LÖHLE, F., 1933, *Nachr. Ges. Wiss. Göttingen, Math. Phys. Kl.*, 271.
- MAYER, J. E., and MALTBIE, M. M., 1932, *Z. Physik*, **75**, 748.
- MASSEY, H. S. W., 1938, *Negative Ions* (Cambridge Physical Tracts), p. 15.
- MEES, C. E. K., 1942, *The Theory of the Photographic Process* (New York: Macmillan), p. 3.
- MITCHELL, J. W., 1948, *Science et Industries Photographiques*, **19**, 361.
- MOTT, N. F., 1937, *Proc. Phys. Soc.*, **49**, extra part, 36; 1941, *Photo. J.*, **81**, 62; 1946, *Journ. de Physique* (8), **7**, 249.
- MOTT, N. F., and GURNEY, R. W., 1940, *Electronic Processes in Ionic Crystals* (Oxford).
- MOLLWO, E., 1933, *Z. Physik*, **85**, 56.
- POHL, R. W., 1937, *Proc. Phys. Soc.*, **49**, extra part, 3.
- ROGENER, H., 1937, *Ann. d. Physik* (5), **29**, 386.
- SAVOSTIANOVA, M., 1934, *Compt. rend. Acad. Sci. U.S.S.R.*, **2**, 228; 1935, *Acta Physiochem. U.S.S.R.*, **111**, 315; 1935, IX Congrès International de Photographie, Paris, 94.
- SHEPPARD, S. E., 1925, *Photo. J.*, **65**, 380.
- SHEPPARD, S. E., and HUDSON, J. H., 1927, *J. Amer. Chem. Soc.*, **49**, 1814.
- SHEPPARD, S. E., TRIVELLI, A. P. H., and WIGHTMAN, E. P., 1927, *Photo. J.*, **281**.
- SIMPSON, J. H., 1949, to be published in *Proc. Roy. Soc. (A)*.
- STASIW, O., and TELTOW, J., 1941, *Nachr. Ges. Wiss. Göttingen, Math. Phys. Kl.*, **93**, 100, 110; 1941, *Ann. d. Physik* (5), **40**, 181; 1944, *Nachr. Ges. Wiss. Göttingen, Math. Phys. Kl.*, 155; 1947, *Ann. d. Physik* (6), **1**, 261; 1948, *Z. Anorg. Chem.*, **257**, 103, 109.
- STROCK, L., 1934, *Z. Phys. Chem. B*, **25**, 441.
- TUBANDT, C., 1932, *Wien-Härms Handbuch d. Experimentalphysik*, Leipzig, **12**, 384.
- WAGNER, C., and BEYER, J., 1936, *Z. Phys. Chem. B*, **32**, 113.

XXIV. *Tank Model for Magnetic Problems of Axial Symmetry.*

By R. E. PEIERLS and T. H. R. SKYRME,

Department of Mathematical Physics, The University, Birmingham, 15*.

[Received September 27, 1948.]

ABSTRACT.

The electrolytic tank can be used to solve magnetic problems involving currents if the arrangement has axial symmetry and the field is required in a region of linear dimensions small compared to the distance from the axis of symmetry, and if the magnetic equipotentials also remain within a region of similar extent.

§ 1.

THE electrolytic tank is commonly used as a model to solve problems of potential theory. In electrostatic problems, which involve fields which are vortex-free, but not necessarily source-free, the tank permits the treatment of spacial as well as plane problems.

Magnetic problems often involve fields which are source-free but not vortex-free. For these a direct representation in the tank is impossible. However, in two dimensions one can utilize the equivalence of conjugate fields to interchange equipotentials and lines of force, and thereby turn a source-free into a vortex-free field (*cf.* Peierls 1946).

This method is limited to the plane case. There is, however, some practical interest in problems of axial symmetry which are very nearly plane, *i. e.* for which the dimension l of the region over which the field extends is small compared to the distance R of this region from the axis. The present note gives a method of treating such problems to first-order accuracy, *i. e.* neglecting the square of l/R .

The solution of the conjugate problem with axial symmetry can be obtained in the tank by setting it up as for the plane problem and then tilting it so that the bottom of the tank and the liquid surface, if extrapolated, would intersect in the axis of symmetry. This does not provide directly a solution of the original magnetic problem for, as we shall see, the curvature effect is of opposite sign in the two problems; however, when that effect is small its value in the magnetic case can be found as the

* Communicated by the Authors.

reverse of the difference between the conjugate solutions obtained in the level and tilted tanks.

Consequently the magnetic field in the curved case may also be obtained by tilting the tank through an equal angle in the opposite direction.

§ 2.

The equations of our magnetic problem are, from Maxwell' equations,

$$\left. \begin{array}{l} \text{div } \mathbf{H}=0 \\ \text{curl } \mathbf{H}=4\pi\mathbf{j}/c \end{array} \right\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where \mathbf{j} is the current density inside any conductors that are present, and is supposed known. Iron, if present, is assumed to have infinite permeability and is represented by the boundary condition :

$$H_{l=0}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

 H_t being the field component tangential to the iron surface.

Taking cylindrical coordinates, r, z, θ , and defining a vector potential A_θ by the relation

$$H_r = -\frac{\partial A_\theta}{\partial z}, \quad H_z = \frac{1}{r} \frac{\partial(rA_\theta)}{\partial r}, \quad \dots \quad (3)$$

we are left with

$$\frac{\partial^2 A_\theta}{\partial z^2} + \frac{\partial}{\partial r} \left[\frac{1}{r} \frac{\partial}{\partial r} (r A_\theta) \right] = -4\pi j/c. \quad (4)$$

Let

$$rA_\theta = f. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Then

$$\frac{\partial^2 f}{\partial z^2} + \frac{\partial^2 f}{\partial r^2} - \frac{1}{r} \frac{\partial f}{\partial r} = -4\pi j r_1 c, \quad (6)$$

One verifies easily that at an iron surface (provided it has axial symmetry) condition (2) amounts to

$$\frac{\partial f}{\partial n_i} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7).$$

n being the direction normal to the surface.

If we neglected the curvature we could represent our problem by a plane tank model with the same cross section, and having current entering the tank with a source density s . The electric potential V in the tank then satisfies the equation (denoting the Cartesian coordinates in the tank by r and z to facilitate comparison)

$$\frac{\partial^2 V}{\partial z^2} + \frac{\partial^2 V}{\partial r^2} = -g/\phi \quad (6A)$$

$$\text{Let} \quad \bar{h} = r H_z = \frac{\partial f}{\partial r}, \quad (17)$$

$$n = \frac{r^2}{\bar{h}} \frac{\partial}{\partial r} \left(\frac{\bar{h}}{r} \right) = \frac{r}{\bar{h}} \frac{\partial \bar{h}}{\partial r} - 1. \quad (18)$$

$$\text{If similarly} \quad \bar{\bar{h}} = \frac{\partial \bar{f}}{\partial r}, \quad h^* = \frac{\partial h^*}{\partial r} \quad (19)$$

$$\text{and} \quad \bar{\bar{n}} = \frac{r}{\bar{\bar{h}}} \frac{\partial \bar{\bar{h}}}{\partial r}, \quad n^* = \frac{r}{h^*} \frac{\partial h^*}{\partial r}, \quad (20)$$

we have, by an easy transformation, and neglecting second-order terms,

$$n = \bar{n} - (n^* - \bar{n}) - 1. \quad (21)$$

$$\text{Or if} \quad h^+ = \frac{\partial f^+}{\partial r} \quad \text{and} \quad n^+ = \frac{r}{h^+} \frac{\partial h^+}{\partial r}, \quad (22)$$

$$\text{then} \quad n = n^+ - 1. \quad (23)$$

REFERENCE.

PIERLS, R. E., 1947, *Nature, Lond.*, **158**, 851.

XXV. Variation with Oxygen Pressure of the Thermoelectric Power of Cadmium Oxide.

By C. A. HOGARTH, Ph.D., and J. P. ANDREWS, D.Sc.,
Queen Mary College*.

[Received August 30, 1948.]

ABSTRACT.

Experiments are described which show that (a) over the temperature range 240°–570° C., the thermoelectric power of a CdO/Pt couple varies with the absolute temperature T according to the equation $dE/dT = -a + b/T$, the pressure of oxygen in the surrounding atmosphere remaining constant, and (b) at constant temperature the thermoelectric power is a linear function of the logarithm of the oxygen pressure. Simple theory predicts this form of variation but not the correct values of the constants. The quantity b varies smoothly with the oxygen pressure as suggested by A. H. Wilson.

1. INTRODUCTION.

THE remarkable dependence of the electrical properties of semiconductors, of which cadmium oxide is an example, upon the nature and number of imperfections in the crystalline structure of the material is now sufficiently

* Communicated by Dr. J. P. Andrews, Queen Mary College.

n_f and σ (if σ can be obtained at all) is simple as a rule. Similar observations apply to the measurement of R , but with the additional objection the R is generally much more difficult to measure. The thermoelectric power ought to be independent of the state of aggregation, but then the connection between dE/dT and n_f is unfortunately not so direct and interpretation is correspondingly more doubtful. Hitherto the conductance of a pressed powder specimen has generally been measured, but in view of the dilemma referred to, it is evidently desirable to supplement the results with measurements of thermoelectric power. Such measurements, however, are not to be regarded as merely supporting or subsidiary, but as having importance in their own right, since some aspects of the behaviour of semiconductors are not so well revealed by conductance measurements.

It may be shown (Hogarth 1948 a) that in a semiconductor whose conduction electrons obey classical statistics, and to which equation (1) may be applied, the thermoelectric power is given as a function of the vapour pressure by the equation

$$dE/dT + \text{constant} = - \frac{k}{e|n} \log P_v, \quad . \quad . \quad . \quad (4)$$

where k is Boltzmann's constant and e is the electronic charge. One or two separate observations by Wagner and his collaborators have been shown to be consistent with this equation (Hogarth 1948 b), but hitherto no systematic investigation has been published.

In the investigation now described, CdO is chosen because its conductivity is large, so that ordinary potentiometer methods can be used in the measurement of dE/dT . The possibility, indeed the probability, that some deviation from classical statistics will ensue from the necessarily high concentration of conduction electrons has to be borne in mind.

2. EXPERIMENTAL ARRANGEMENT.

Specimens of cadmium oxide 8 cm long, 1.75 cm in diameter, made by compression from the purest obtainable CdO powder, were used, the powder having been baked for two hours at 700° C. before pressing. A specimen of this kind, housed in a fused silica tube, was held firmly between thin platinum electrodes from which platinum leads were brought out through fused-on glass capillaries—in order to give rigidity and make a gas-tight system. To measure the temperature at each end of the specimen, chromel-alumel thermojunctions were silver-soldered to the electrode disks. The complete cell having been made gas-tight, connection was made to a pumping system and pressure gauges. A water jacket at one end of the test cell and a small electric furnace at the other produced a temperature difference between the ends of the specimen, the water-jacketed end being maintained at a few degrees above room temperature throughout. The electric circuit included a standard potentiometer arrangement with its appropriate controls (see fig. 1).

Fig. 1.

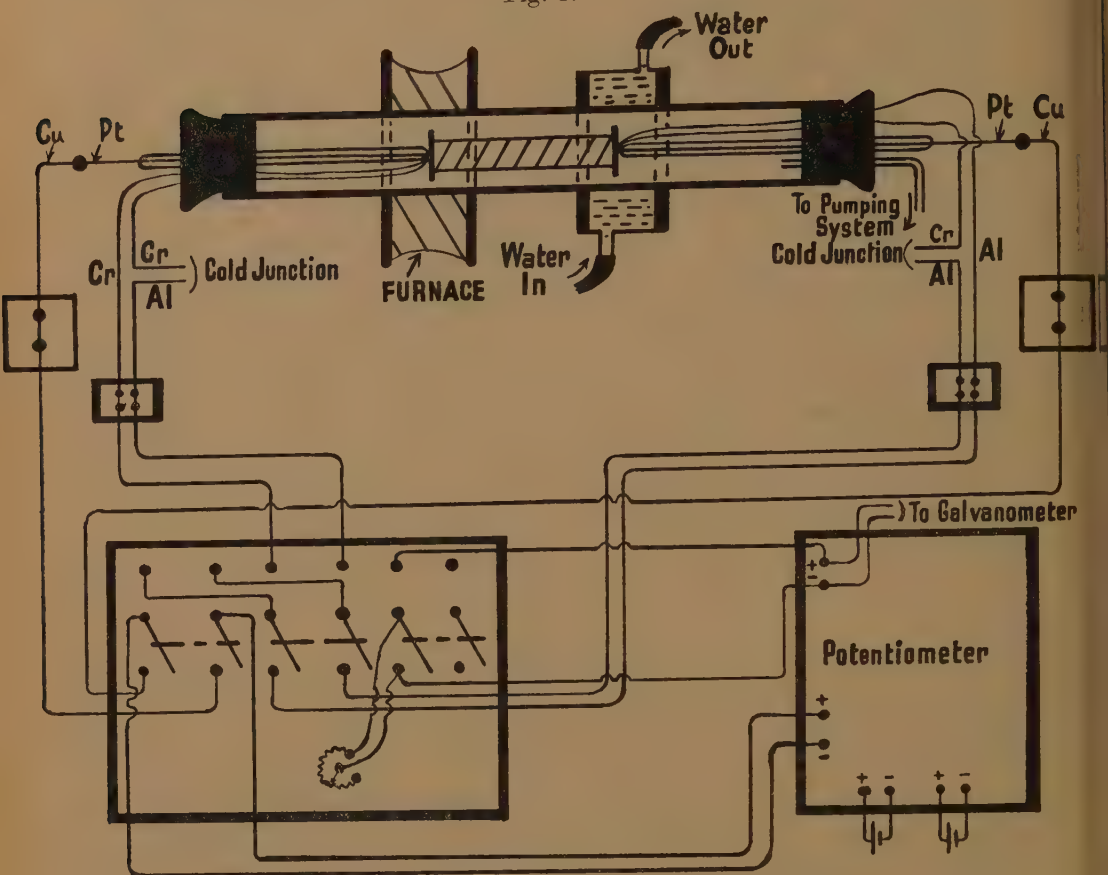


Diagram of test-cell and associated electrical circuit.

3. PROCEDURE.

The system having been evacuated to a pressure less than 0.1 mm. of mercury, dry air or oxygen was introduced as required, low oxygen pressures being obtained by first filling with oxygen at atmospheric pressure and then pumping out to the pressure required. The furnace was then switched on, the temperature rise being arrested at intervals of approximately 15° for the measurement of the thermo e.m.f. of the Pt/CdO/Pt system. Heating was continued in this way until the temperature of the hot junction reached about 570°C . Dry air at atmospheric pressure was then admitted and the furnace switched off. By adhering to this procedure it was possible to get reproducible values of the e.m.f. over a period of weeks, and the moderate heat treatment did not affect the electrical characteristics of the specimen permanently. The oxygen pressure in the system could be held constant for an indefinite time.

From the slopes of a series of graphs of the thermo e.m.f. against the difference of temperature between the ends of the specimen, the thermo-electric power dE/dT as a function of T , the temperature of the hot end

was obtained. During the experiment it was found impossible to prevent the temperature of the cold end from changing by a few degrees. A calculation of the error in dE/dT due to this cause showed that it did not exceed one per cent in the most serious case, and it was considered safe to neglect any correction of dE/dT on this account. Values of dE/dT thus obtained were found to be given by the empirical equation

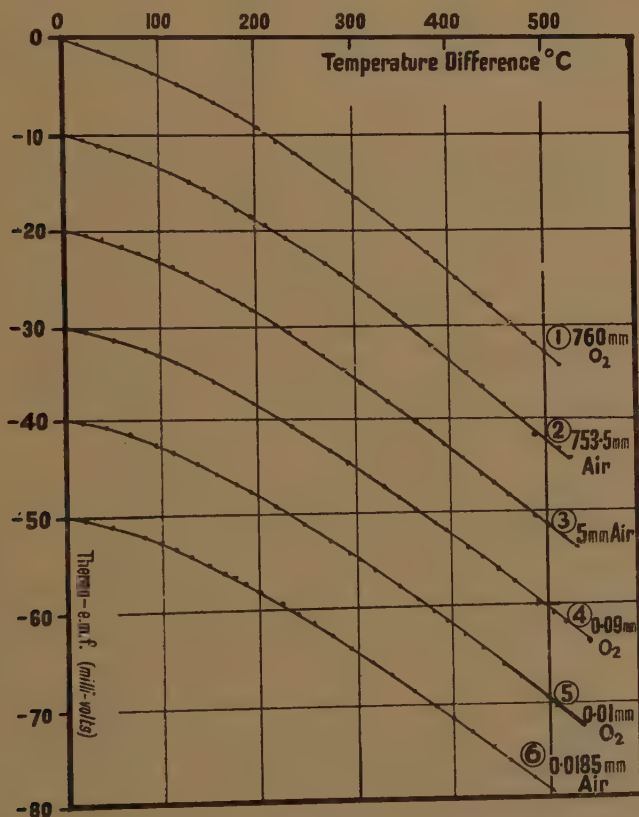
$$dE/dT = -a + b/T, \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where a and b are constant at any one pressure, and over a defined temperature range.

4. RESULTS OF EXPERIMENTS.

Experiments were performed on three specimens of CdO prepared as described. Their densities were 3.9, 4.2 and 4.5 gm. cm.⁻³, or about half the crystal density of CdO, and all three specimens behaved very similarly.

Fig. 2.



Thermo e.m.f. and temperature difference at 6 different pressures of oxygen. (The origins of successive curves have been displaced in steps of 10 milli-volts. The scale is corrected for curve 1; for curve 2 add 10 mv. and so on.)

The results about to be quoted apply to the third of these specimens. In fig. 2 the variation of e.m.f. with temperature difference is shown for

six different air or oxygen pressures, while in fig. 3 dE/dT is plotted against $1/T$. It is seen that dE/dT is, substantially, a linear function of $1/T$. At any one temperature the numerical value of dE/dT changes in the

TABLE I.
Data relating to Fig. 3 and Equation (5).

Graph no.	Pressure (mm. of Hg)	Partial press. of oxygen, P_v (mm. of Hg)	$\log_{10} P_v$	$\frac{a}{\text{volt } ^\circ\text{C.} \times 10^3}$	$\frac{b}{\text{volt} \times 10^3}$
1	760, O ₂	760	2.880	0.130	30.35
2	754.5, Air	150.7	2.178	0.126	28.36
3	5.0, Air	1.0	0.0	0.118	20.66
4	0.090, O ₂	0.090	2.954	0.099	16.57
5	0.010, O ₂	0.010	2.000	0.094	14.79
6	0.0185, Air	0.0037	3.568	0.089	12.15

TABLE II.
Values of C_T in Equation (6).

$\frac{10^4}{T ^\circ\text{K.}}$	$T ^\circ\text{K.}$	C_T (e.m.u.)
12	833	369
13	769	335
14	714	300
15	667	266
16	625	232
17	588	196
18	556	164
19	526	130

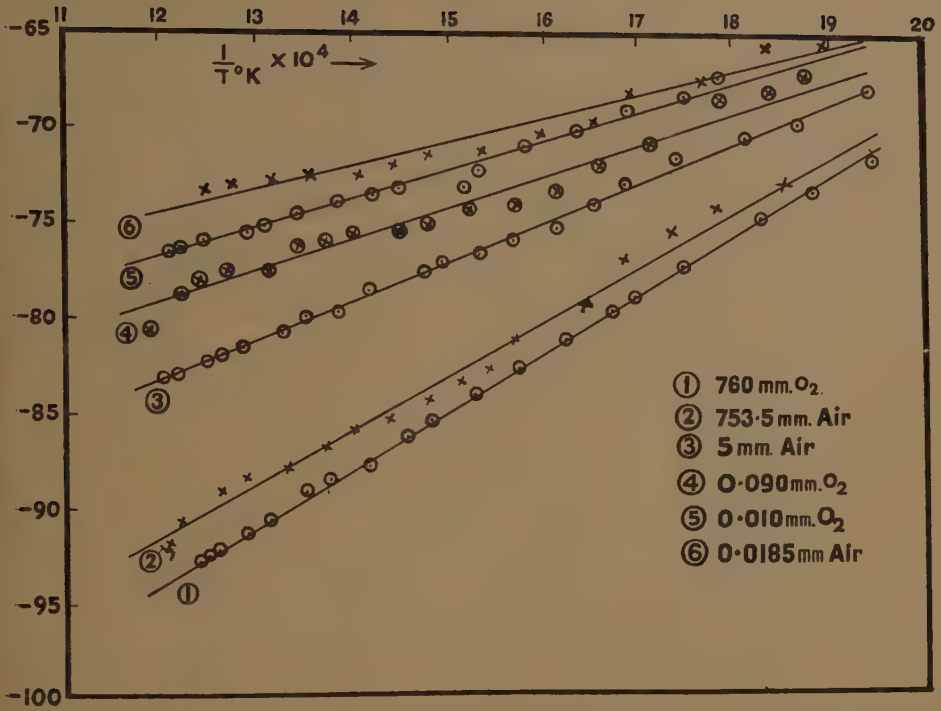
same sense as the oxygen pressure, and fig. 4 shows that dE/dT is a linear function of the logarithm of the oxygen pressure when the temperature is kept constant, or

$$dE/dT = (C_T \log P + \text{const.}) \quad (6)$$

The quantity C_T is a function of the temperature as fig. 5 shows.

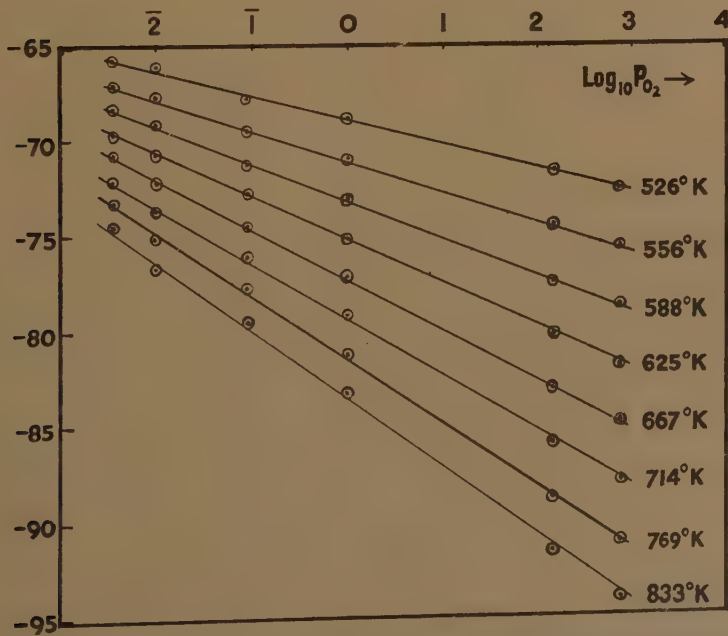
Values of the constants a and b , and of C are given in Tables I. and II.

Fig. 3.



The variation of thermoelectric power of cadmium oxide with temperature for six different pressures.

Fig 4.



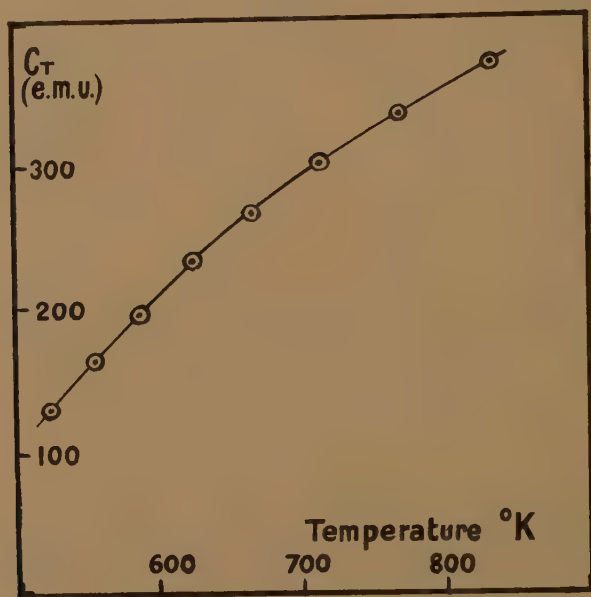
The variation of thermoelectric power of cadmium oxide with oxygen pressure.

From the graphs it may seen that

(a) An atmosphere of air has the same effect as an atmosphere of oxygen whose pressure is equal to the partial pressure of oxygen in the air, so that nitrogen has no measurable effect. This agrees with the observations of von Baumbach and Wagner on conductivity.

(b) The variation of thermoelectric power with temperature, the oxygen pressure being constant, follows the same general law for all oxygen pressures within this range, that is, between 4×10^{-3} and 760 mm. pressure, and from 240° to 570° C.

Fig. 5.



The variation of C_T with temperature.

For all specimens it was found that a change in gas pressure was followed by an almost instantaneous change in the e.m.f. and in all cases equilibrium appeared to be established within two minutes. After the attainment of equilibrium, provided the temperature and oxygen pressure were kept constant, the e.m.f. remained unchanged for periods up to 8 hours. The effects were in all cases reversible, and observations on a specimen gave the same results over a period of weeks.

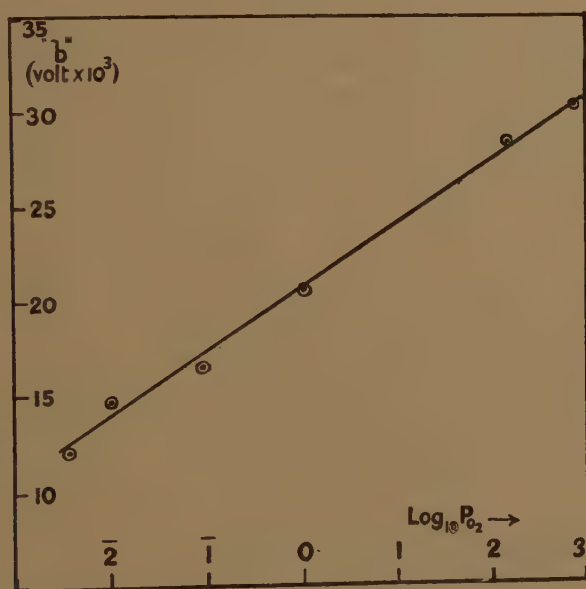
A few rough measurements of resistance indicated that a change of oxygen pressure altered the conductance in the direction to be expected in an excess conductor - pressure and resistance decreased together.

5. DISCUSSION OF RESULTS.

Perhaps the most interesting result is the parallel expressed by equations (4) and (6) between the experimental results and the simplified theory. The resemblance suggests an obvious interpretation of the experimental

results, the kind of interpretation not infrequently made in work on solids; and indeed it does seem likely that the simple theory is on the right lines. However, the parallel extends only to the outward form of the relation. Thus if the two equations are identified for the moment, n can be calculated from C_T . In this way values of n ranging from 50 at 500°C. to 150 at 200°C. are found. It is difficult to imagine a mechanism which would yield anything like these values. Starting with the mechanism described in the Introduction, and making the same kind of supplementary assumptions as were made by Wagner in his study of the similar compound ZnO , the value $n=6$ is obtained, and it is worth recalling that experiments on the conductivity of ZnO at different oxygen pressure are consistent with this value of n . CdO , however, has a much

Fig. 6.



The variation of "b" with oxygen pressure.

larger conductivity than ZnO , and a much denser concentration of free electrons may be assumed, with a probable departure from the Maxwellian distribution of velocities. Moreover, since thermoelectric effects depend on the first power of the carrier charge, while conductivity depends on its square, a further complication must ensue if carriers of both signs contribute. The exceptionally low value of dE/dT in CdO does suggest the presence of both kinds of charge. In any adequate theory of the thermoelectric effect in CdO both these complications must be taken into account. In one other particular the results suggest the simpler theory. A. H. Wilson (1939) has remarked that the value of the constant b in equation (5) should vary smoothly with the vapour pressure in any semiconductor which obeys this equation and which is sensitive to changes of the pressure. Fig. 6 shows that b is a linear function of $\log P_v$.

CONCLUSIONS.

1. The thermoelectric power of CdO at constant temperature is a linear function of the logarithm of the surrounding oxygen pressure (equation (6)).

2. By an application of the free-electron theory of metals and the "Fehlordnungstheorie" of Wagner and Schottky, a relation of the same form can be deduced; but the correspondence does not extend to numerical magnitudes and it is unlikely that the simple theory can be applied without modification to CdO.

3. If the surrounding atmosphere is air, the effective vapour pressure is the oxygen partial pressure.

4. Between 240° C. and 570° C. the variation of thermoelectric power with temperature at constant oxygen pressure is represented by the equation $dE/dT = -a + b/T$. The equation has the same form for all pressures between 10⁻³ and 760 mm., the constants a and b being functions of the pressure.

5. The effects of change of oxygen pressure are reversible.

6. After the attainment of equilibrium the thermoelectric power remains constant for at least 8 hours.

7. The quantity b in the equation $dE/dT = -a + b/T$ is a linear function of $\log P_v$.

ACKNOWLEDGMENTS.

These experiments were carried out with the assistance of a grant from the Department of Scientific and Industrial Research. We should like to record our thanks to Prof. H. R. Robinson for providing us with facilities in his laboratory. The substance of this work formed part of a Thesis submitted for the degree of Ph.D. in the University of London.

REFERENCES.

- ANDREWS, J. P., 1947, *Proc. Phys. Soc.*, **69**, 990.
 VON BAUMBACH, H. H., and WAGNER, C., 1933, *Zeits. Phys. Chem. B*, **22**, 199.
 HOGARTH, C. A., 1948 a, *Phil Mag.* ser. 7, **39**, 260; 1948 b, *Nature, Lond.*, **161**, 60
 WILSON, A. H., 1939, *Semiconductors and Metals*, p. 49 (Cambridge).

XXVI. *The Electrical Conductivity of Bismuth Fibres.—I. Magneto Resistance and the Crystalline Structure.*

By B. DONOVAN and G. K. T. CONN,
Department of Physics, The University of Sheffield *.

[Received November 15, 1948.]

INTRODUCTION.

THE exceptional galvano-magnetic effects exhibited by bismuth have been the subject of extensive study for at least 50 years, although it is only of late that any marked clarification has been possible. In particular, the change in electrical resistance due to a magnetic field is very much larger in the case of bismuth than for any other element. Specimens in the form of bismuth spirals have been used as a convenient, rapid means of measuring magnetic field strength.

The early work on polycrystalline specimens has been discussed by Campbell (1923). At optical and infra-red frequencies the influence of a magnetic field on conductivity is negligible (Heaps 1926, McLennan *et al.* 1932). Blunt (1947) has reported that the behaviour in a magnetic field is the same using alternating currents of frequencies up to 3.5 megacycles per second as it is using direct current. In recent years emphasis has shifted to single crystals and there is now a considerable amount of information on such specimens. Kapitza (1928) made an extensive study of the magneto-resistance effect using very high fields of short duration. Schubnikow and de Haas (1930) investigated the behaviour at temperatures down to 14° K. and de Haas, Blom and Schubnikow (1935) found that the effect was even more pronounced at liquid helium temperatures. The importance of the relative orientation of the crystal axes, the current and the magnetic field, was soon realized, and detailed information was obtained by Stierstadt (1933, 1934, 1935), Grüneisen and Gielessen (1936, 1937) and Kaye (1939). The significance of impurities in the bismuth was investigated by Thompson (1938), who confined his measurements to certain specific orientations of the crystal. The most important work from the point of view of the present paper is that of Stierstadt, because of the comprehensive nature of his investigations. In his six papers (subsequently referred to as Stierstadt a, b, etc., in chronological order) the resistance changes are studied in detail and interpreted in terms of the crystal structure.

* Communicated by Prof. W. Sucksmith, D.Sc., F.R.S.

Interpretation of such results has been furnished in terms of the classical theory of anisotropic media (Kohler 1934). The state of development of the electron theory of metals is, however, not yet adequate. The approximate quantitative treatment due to Jones (1936) assumes circular symmetry round the principal axis, which in fact is a trigonal axis, whilst a more recent paper by Sondheimer and Wilson (1947) goes no further in this direction.

The single crystals which hitherto have been used in magneto-resistance investigations suffer from one inherent disadvantage. The mode of preparation of these specimens is such that the lateral diameter must be at least of the order of a millimetre, and as a result the electrical resistance is low. In consequence the finer details of the electrical behaviour are necessarily elusive.

The purpose of the present paper is to set out the results of the first of a series of investigations using very thin fibres of bismuth. By drawing down in a soda glass envelope it is a simple matter to prepare fibres whose resistance is of the order of tens or even hundreds of ohms per centimetre (see Table I.), and the relative reliability offers an easy means of investigating the behaviour of the metal. It might be expected, on account of the mode of preparation, that the fibres should approximate in some degree to single crystals, and this expectation is, in fact, borne out by the results of the magneto-resistance experiments to be described. Because of the comparatively high resistance, simple use of a potentiometer enables close study to be made of single crystals prepared by these methods.

So far as the authors are aware, high resistance specimens of this type have not hitherto been used to examine the magneto-resistance properties of metals. An examination of the literature shows that Eucken and Förster (1934) claim to have determined a value for the mean free path of the conduction electrons in bismuth from an observed anomaly in the resistivity-temperature curve of such fibres.

In the present paper attention is limited to a study of the crystalline state by examining the change of resistance in a magnetic field as the orientation of the fibre is changed. A brief note on these experiments has already been published (Conn and Donovan 1948). Further reports, including the evidence obtained by X-ray diffraction methods, will be presented later.

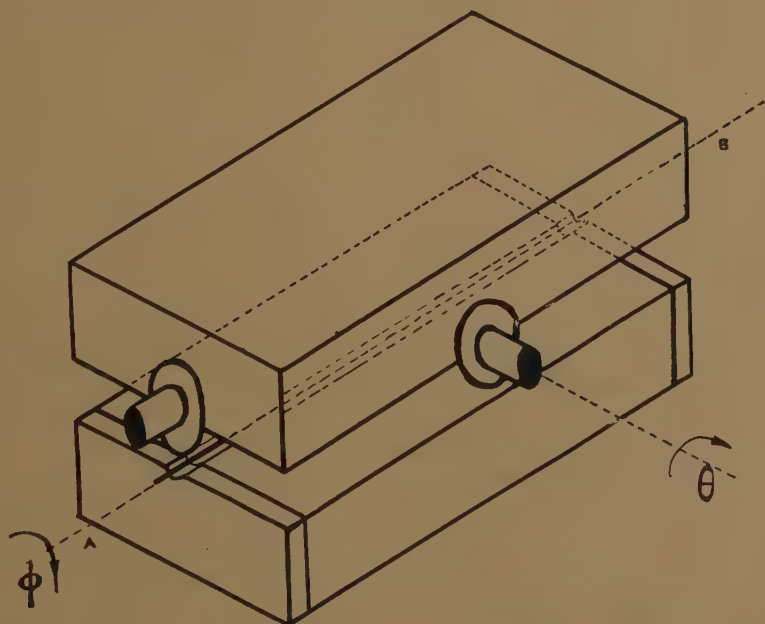
EXPERIMENTAL.

The metal used was commercial bismuth with a measured impurity content of less than 0.1 per cent. The bulk of this impurity was lead, and as the emphasis in these experiments was not primarily on the quantitative results spectroscopically pure material was not used.

The specimens were prepared by hot-drawing in soda glass tubing, and were of the order of 5–150 μ in diameter. The drawing process was carried out in two stages, firstly in a Bunsen flame and secondly in a nichrome heater coil. It is important to be able to control the temperature of this coil; by raising the temperature the rate of drawing can be markedly

increased and much finer fibres obtained. The bismuth was drawn down in the soda glass and solidified under pressure because of the increase of about 3 per cent in the volume accompanying the change of state. Lengths were cut of approximately 2 cm., the glass etched from the ends with hydrofluoric acid and the fibres soldered to narrow copper strips mounted on an ebonite holder (fig. 1). Several low melting point solders were tried, but Wood's metal, an alloy containing a high proportion of bismuth (Strong 1940, p. 548), was most satisfactory. Resin dissolved in alcohol was used as a flux. A horizontal brass tube carried the specimen

Fig. 1.



Isometric drawing of the holder. AB is the axis of the fibre and the ϕ -axis.

holder and supported it between the pole pieces of an electro-magnet. The tube, which was perpendicular to the magnetic field, provided the axis of rotation of the specimen; angular displacement could be read on a scale at the end of the tube. The specimen holder could be mounted so that the fibre axis was either parallel to, or perpendicular to, the axis of rotation. Two orientations were used for reasons of geometric adequacy.

By means of a potentiometer, the resistance was compared in the usual way with a standard resistance. In many cases, particularly where the resistance changes were small, it was convenient to use the galvanometer as a continuous recorder of changes of potential difference. Since changes of the order of 10^{-6} volt could be detected, small temperature fluctuations in the specimen were frequently responsible for a slow drift of the galvanometer spot, but this could usually be eliminated by sliding a test-tube over

the specimen-holder and packing the end with cotton wool. To avoid possible complications due to transverse effects the current was reversed in all cases and all sets of readings were duplicated. Every resistance "cycle" illustrated in this paper is the mean of two complete curves, one for each direction of the current. In general, any discrepancy in the values of ΔR for the two directions of the current was less than 2 per cent. Certain important anomalies will be discussed elsewhere.

Before presenting the results obtained, a brief comment on the orders of magnitude is desirable. An estimate of the diameter of the fibres used may be obtained using the value of the resistivity perpendicular to the principal axis, which is 114×10^{-6} ohm-cm. at 25°C . (Kaye 1939). The reason for choosing this value of the resistivity will appear later. Taking a mean length of 1.7 cm., diameters are listed in Table I. for specimens of various resistance.

TABLE I.

Resistance in ohms	1	50	160	800
Diameter in μ	160	22	12	5

The technique of etching and mounting the fibres raises the possibility that a contact resistance (or e.m.f.) may reduce the apparent value of the ratio $\Delta R/R$, where ΔR is the resistance change due to the magnetic field. The danger is not serious in practice, since inadequate etching is easily observed under low magnification and leads to a resistance value which is obviously absurd. Moreover, the accuracy with which resistance is measured is such that anything in the nature of an erratic contact cannot be tolerated. The matter was, however, further tested, since fibres of bismuth, prepared by a technique similar to ours, have been used by Eucken and Förster (1934) in an attempt to estimate the mean free path of electrons in bismuth. By measuring the resistance of a number of fibres at 0°C . and at -72°C ., the temperature coefficient was determined and found to be about 3.0×10^{-3} per $^\circ \text{C}$. While the temperature coefficient of resistance of bismuth crystals warrants further investigation (Pietenpol and Miley 1929, Eucken and Förster, *loc. cit.*), these measurements make it clear that the method of mounting the specimens was satisfactory. This was confirmed in a number of cases by direct measurements of the diameter.

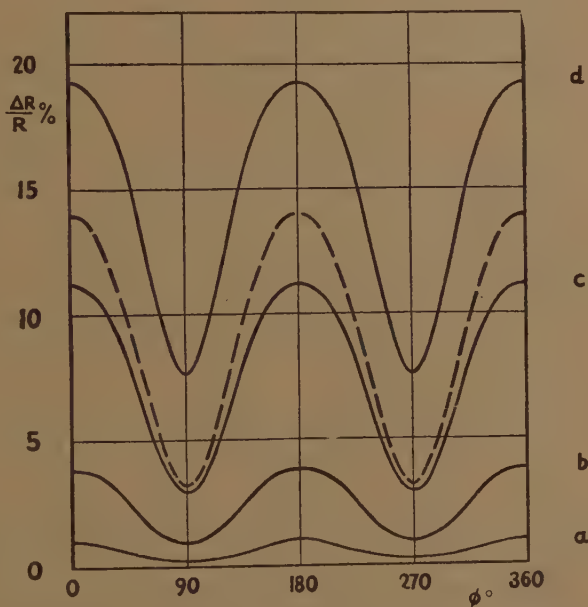
VARIATION OF RESISTANCE WITH ϕ .

In the first series of results to be presented the axis of the specimen coincided with the axis of rotation and was therefore perpendicular to the magnetic field. Thus the direction of the current in the specimen was always perpendicular to the field. The percentage increase in resistance

due to the magnetic field was plotted as a function of the angle of rotation which, for this orientation, has been denoted by ϕ . About 50 different fibres were examined, the resistances ranging from 1 to 800 ohms.

For the majority of the fibres the same form of resistance variation with ϕ was obtained, and in view of this uniformity only one series of resistance cycles need be given. The exceptions were the fibres with the lowest values of resistance, and these will be discussed later. Fig. 2 shows representative results obtained at four different field strengths with a fibre whose resistance, in the absence of the field, was 42 ohms. These curves may be regarded as typical. Absolute values of $\Delta R/R$ obtained

Fig. 2.



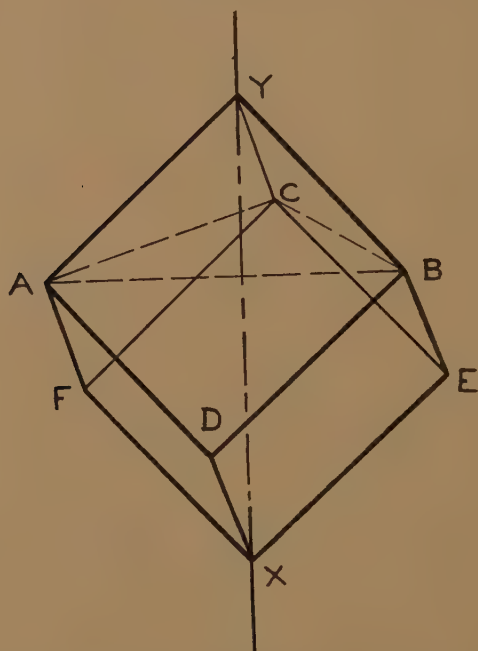
Variation of resistance with ϕ for 42 ohm fibre. The field strengths, in oersteds, for the four curves are (a) 990, (b) 2290, (c) 4400 and (d) 6300. The broken line represents Stierstadt's result for a P_1 (\perp) single crystal at 4400 oersteds.

with different fibres at one particular field strength must be compared at corresponding orientations relative to the field, for instance at the maximum or the minimum of the curves. Such absolute values were consistent in all cases within 10 per cent, although no precautions were taken to ensure that the temperature was constant from day to day.

The main features are apparent from fig. 2. The curves have the same periodic character with a periodicity of π . The amplitude increases with increasing field strength although the form of the curve does not change appreciably. It should also be noted that the maxima are somewhat broader than the minima.

These curves may be compared with similar curves obtained using single crystals of bismuth. Fig. 3 shows the single crystal of bismuth. XY is the principal, trigonal axis and the main cleavage plane (111) is perpendicular to this direction. The three subsidiary cleavage planes of the form $(11\bar{1})$ are parallel to the faces $YADB$, $YBEC$, and $YCFA^*$. These three planes intersect the main cleavage plane in AB , BC and CA respectively and these three directions are parallel to the three binary axes.

Fig. 3.



Bismuth single crystal. XY is the principal axis and the three binary axes are parallel to AB , BC and CA .

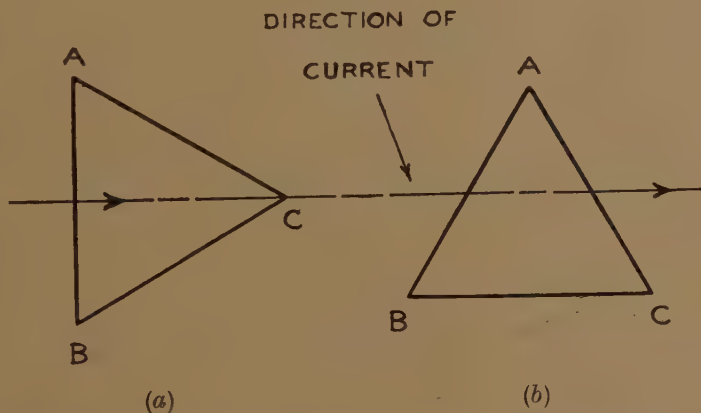
In his second paper, Stierstadt (1933 b) reports resistance changes of a single crystal for which the direction of current-flow was always parallel to the axis of rotation and perpendicular to the magnetic field. The most illuminating curves are those obtained with P_1 type crystals (1933 b, figs. 7 and 10). The notation is that of Stierstadt, and in these crystals the principal axis is perpendicular to the axis of rotation and accordingly the main cleavage plane is parallel to the direction of the current. Two important cases arise according as the current is perpendicular $P_1 (\perp)$, or parallel $P_1 (\parallel)$ to one of the binary axes. These are illustrated in fig. 4. The curves for the $P_1 (\perp)$ strongly resemble our fig. 2 at low field strengths.

* The statement of Kaye (1939) that the subsidiary cleavage planes are orthogonal to the main cleavage plane is incorrect.

The periodicity of π is observed in all cases and the resistance changes are of the same order of magnitude. For comparison the curve obtained by Stierstadt (1933 b, fig. 7) for a $P_1(\perp)$ crystal at 4400 oersteds (3500 AW/cm.) is reproduced in fig. 2. The maximum in our case is somewhat smaller. There is no curve for a corresponding field strength in fig. 10 (Stierstadt 1933 b), but a rough estimate shows that the values here are closer to our fig. 2.

Such comparisons suggest forcibly that these thin filaments are single crystals in which the principal axis is perpendicular to the axis of the specimens. The absolute values of $\Delta R/R$ make it clear that the approximation to a single crystal must be very close. Fibres in which the direction of the binary axis AB varies between the positions (a) and (b) of fig. 4 should give curves of slightly different amplitudes. Rough quantitative

Fig. 4.



Orientation of binary axes with respect to current in (a) $P_1(\perp)$ crystal and (b) $P_1(\parallel)$ crystal. Main cleavage plane parallel to paper.

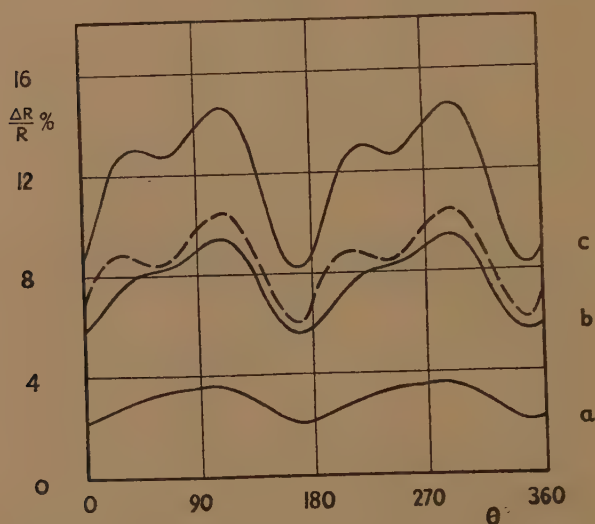
comparison indicates that the crystalline type is intermediate between $P(\perp)$ and $P(\parallel)$, though in general nearer the latter. The observed spread of the order of 10 per cent in the values of $\Delta R/R$ referred to above is presumably due to difference in the direction of the binary axis AB.

If the principal axis were not perpendicular to the axis of the specimen, fig. 2 would not show a periodicity of π . If, for instance, the principal axis were parallel to the axis of the specimen the resistance changes due to rotation would show a six-fold symmetry. Such a crystal-type (P_3) has been shown by Stierstadt (1933 b, fig. 13) to give the expected periodicity of $\pi/3$. It is difficult to see how regular curves having the form and periodicity of fig. 2 could be obtained consistently unless the principal axis is perpendicular to the axis of the specimen. This conclusion is to be anticipated from the method of preparation. In drawing down a fibre one would expect slip to occur along the main cleavage planes. It is reasonable, therefore, that the main cleavage plane is parallel to the axis.

VARIATION OF RESISTANCE WITH θ .

In this series of measurements the axis of rotation was again perpendicular to the field but also perpendicular to the fibre, passing approximately through its centre. The fibre was thus rotated in a plane and the angle of rotation gave the angle between the fibre, and hence the current, and the field. This is illustrated in fig. 1. For this orientation the angle of rotation has been denoted by θ . At 0° and 180° the fibre was parallel to the field; at 90° and 270° it was perpendicular to the field. The first distinction between the geometric arrangement in the two cases is that in the second there is a natural reference position from which θ may be

Fig. 5.



Variation of resistance with θ for 42 ohm fibre. The field strengths, in oersteds, for the three curves are (a) 2200, (b) 3750 and (c) 5250. The broken line represents Stierstadt's result for a P_1 (intermediate) single crystal at 3700 oersteds.

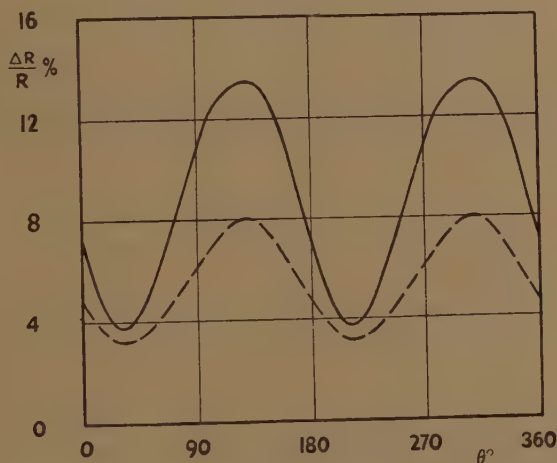
measured, namely, when the fibre is parallel to the field. There is of course no such convenient reference line in the case of variation with respect to ϕ . The second distinction follows from geometric considerations. The fibre is anisotropic and therefore variations about three axes must be studied to designate the crystalline state completely. The third axis may be considered as that axis perpendicular both to the ϕ -axis and to the θ -axis. It is thus in effect a second θ -axis perpendicular to the first. The problem of discriminating θ -axes is that of designating a zero in the case of the ϕ -variation. Since there is no means, a priori, of so doing, no effort was made to obtain two θ -curves for a given specimen by investigating the behaviour about two perpendicular axes. The

properties of any particular fibre were in the present investigation subsidiary to the general properties of fibres prepared by this means. The various possibilities appear in studies of different fibres. If for any reason a specific designation of the crystalline state is sought of a particular fibre prepared in the way described, it would be necessary to investigate the variation of resistance about three axes.

The emergence of such significant conclusions from the first set of measurements is of material assistance in the interpretation of the observed variation about the θ -axis. As is to be expected, curves giving the fractional increase in resistance as a function of θ showed some diversity. One important characteristic, namely, a periodicity of π was common to all.

Fig. 5 illustrates a variation frequently encountered. This shows curves obtained at three different field strengths with the fibre used in the

Fig. 6.

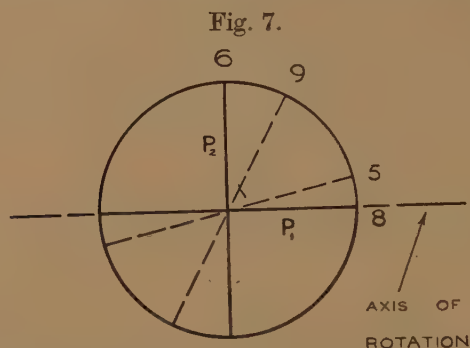


Variation of resistance with θ for 780 ohm fibre at 5250 oersteds. The broken line represents Stierstadt's result for a $P_2(\perp)$ single crystal at 3750 oersteds.

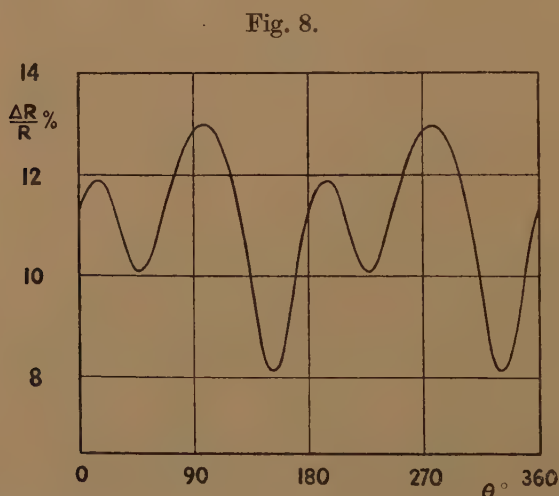
case of fig. 2. The lack of symmetry is immediately apparent, and the shape of the curve alters as the field is increased although the periodicity is constant. The peaks show progressive distortion, and eventually a subsidiary maximum appears accompanied by displacement of the main maximum.

The other common type of θ -variation is shown in fig. 6 which was obtained with a fibre of resistance 780 ohms. The curve is almost symmetrical in that the maxima fall very nearly midway between the minima. No appreciable change in shape was obtained up to the maximum field available to us. This maximum field was about 10,000 oersteds when the pole pieces are set to accommodate variation with respect to θ . It should be further remarked that the amplitude of the variation is considerably greater than that of the corresponding curve of fig. 5, that for 5250 oersteds.

We may assume from the previous discussion that we are dealing with a crystal in which the principal axis lies in a plane perpendicular to the axis of the fibre but is otherwise unknown. Reference may be made to the first paper of Stierstadt (1933a), for the relevant curves obtained with a single crystal. If the principal axis is parallel to the axis of rotation and therefore remains perpendicular to the field throughout the cycle, we have



Orientation of principal axis with respect to rotation axis in P_1 and P_2 crystals. The numbers refer to other figures in the text, which apply to specimens having the principal axis in the direction indicated. The dotted lines are conjectural.



Variation of resistance with θ for 160 ohm fibre at 5250 oersteds.

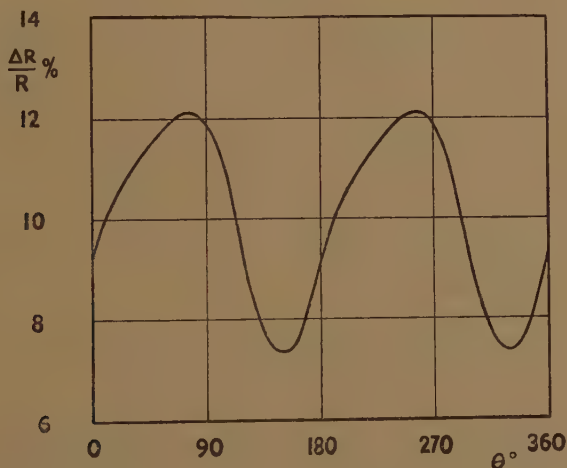
the type designated P_1 by Stierstadt (1933a). Further subdivision is possible if one of the binary axes is parallel, P_1 (\parallel), or perpendicular, P_1 (\perp) to the direction of the current. These subdivisions are illustrated in fig. 4. If the principal axis of the crystal is perpendicular to the axis of rotation, we have the case designated P_2 (Stierstadt 1933a). Fig. 7 shows a section through the fibre perpendicular to the length.

The curves in fig. 5 show that this specimen must closely resemble a P_1 crystal, intermediate between the cases designated P_1 (\perp) and P_1 (\parallel).

For comparison the curve obtained by Stierstadt (1933a, fig. 8) for such an intermediate crystal at 3000 AW/cm. is reproduced in fig. 5*. The subsidiary maximum found by Stierstadt is less pronounced in our curve at the same field strength of 3750 oersteds but the numerical agreement is reasonably good. It therefore appears that this specimen is not a perfect P_1 crystal but that the principal axis is inclined at a small angle to the axis of rotation (see fig. 7).

The form of variation illustrated in fig. 6 is exactly similar to that shown by P_2 (1) crystals and the numerical agreement is close. Stierstadt's curve (1933a, fig. 13) for a P_2 (1) crystal at 3000 AW/cm. is reproduced in fig. 6. For 5250 oersteds the maximum and minimum values of $\Delta R/R$, estimated from Stierstadt's data, are 14 and 5 per cent respectively, while the values in fig. 6 are 13.4 and 3.7 per cent. It would seem very probable that in this instance the specimen is a true P_2 (1) crystal.

Fig. 9.

Variation of resistance with θ for 50 ohm fibre at 5250 oersteds.

The other fibres investigated show, with few exceptions, curves which differ but slightly from the above patterns. The most common is that illustrated in fig. 6. For this reason only two other examples need be given. Fig. 8 is a curve obtained at 5250 oersteds with a fibre of resistance 160 ohms. It demonstrates in a more extreme form the phenomenon of the double maximum shown in fig. 5. In this case the pronounced nature of the subsidiary maximum suggests that the orientation approaches closely that of P_1 types. It is interesting to record that no fibre has given an extreme P_1 (1) curve (Stierstadt 1933a, fig. 5) or a P_1 (||) curve (Stierstadt 1933a, fig. 7).

Finally, as an example of a crystal intermediate between the types P_1 and P_2 we may cite fig. 9. This was obtained with a fibre of 50 ohms

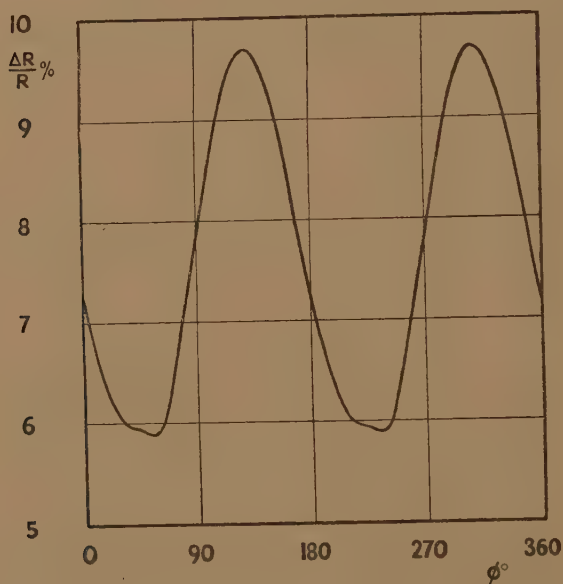
* The apparent lateral inversion of Stierstadt's fig. 8 when compared with our fig. 5 arises because positive rotation in his case corresponds to negative rotation in ours.

resistance and a field strength of 5250 oersteds as in figs. 6 and 8. The characteristic double-hump of a P_1 crystal is absent although the peak is not symmetrical. The amplitude is too small for a genuine P_2 crystal and it seems reasonable to suppose that the principal axis has an intermediate orientation as indicated by the dotted line (9) in fig. 7.

ANOMALOUS FIBRES.

In an earlier section reference was made to the fact that the behaviour of low resistance fibres did not strictly conform to the general behaviour.

Fig. 10.



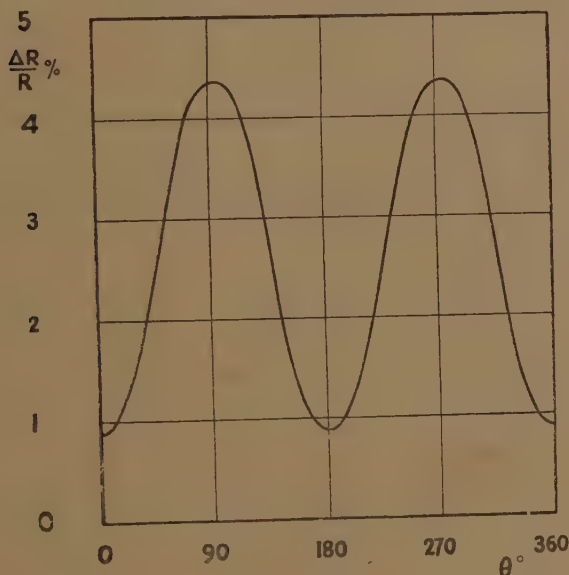
Variation of resistance with ϕ for 1 ohm fibre at 5250 oersteds.

In view of the deduced crystalline of the specimens this divergence does not appear unnatural and might be anticipated. The fibres with the lowest resistance have the greatest area of cross-section and in these cases it is to be expected that the crystalline formation would not be so regular.

Two examples illustrate this supposition. The first, fig. 10, is a ϕ -variation obtained at 5250 oersteds with a fibre of resistance 1 ohm. The similarity to fig. 2 in general shape and periodicity is well-marked, so that even this comparatively thick fibre shows preferred orientations. However, the maximum value of $\Delta R/R$ is even less than that of the standard curve at 4400 oersteds (fig. 2), and the amplitude is only half as great. This suggests that the crystal is not highly developed and irregularities would account for the shape of the curve in the neighbourhood of the minima.

The second example using a fibre of resistance 16 ohms shows in fig. 11 the variation with respect to θ taken at the same field strength as in fig. 10. The form of the curve suggests a P_2 crystal, but the peak value of $\Delta R/R$ and the amplitude are much less than is to be expected with either a genuine P_2 (\perp) or P_2 (\parallel) type. Again we may conclude that the crystal development is incomplete although there is a sufficient degree of preferred orientation to show the characteristic shape of the curve.

Fig. 11.



Variation of resistance with θ for 16 ohm fibre at 5250 oersteds.

CONCLUSION.

From the foregoing experimental results we may assert with some confidence that the fibres, prepared in the manner described, are of a highly ordered structure and, if the resistance be not too low, approximate to single crystals. That the preferred orientation of the cleavage plane is approximately parallel to the specimen axis seems conclusive from the measurement of the variation of resistance with orientation in a magnetic field. Indeed the ϕ and θ curves of all fibres exhibit a periodicity of π . Any departure of the principal axis from a plane perpendicular to the specimen axis would make itself apparent by a change in the periodicity which, therefore, may be regarded as a test of some sensitivity. Of the orientation of the subsidiary cleavage planes the measurements of magneto-resistance are not so conclusive though much significant information can be obtained.

SUMMARY.

Experiments have been carried out on the change of electrical resistance of bismuth in a magnetic field using specimens in the form of thin fibres. Such specimens, prepared by the Taylor process, are superior to previous single crystals in respect of high resistance (~ 10 – 100 ohms) and ease of preparation. The variation of resistance in a magnetic field as a function of orientation has been studied in two main cases. In one case (ϕ variation) the fibre axis is parallel to the axis of rotation and in the other (θ variation) it is perpendicular to the axis of rotation; in both cases the axis of rotation is perpendicular to the magnetic field. By detailed comparison of the resistance changes with those obtained using single crystals much information has been obtained on the crystalline state of the fibre. In particular, the conclusion emerges that, provided the resistance is not too low the fibres approximate closely to single crystals with the main cleavage plane parallel to the fibre axis. For thicker fibres, having resistances below about 15 ohms/cm., the probability of obtaining a single crystal is small, although there is much evidence of preferred orientations.

REFERENCES.

- BLUNT, R. F., 1948, *Phys. Rev.*, **73**, 654.
 CAMPBELL, L. L., 1923, *Galvanomagnetic and Thermomagnetic Effects* (London: Longmans Green).
 CONN, G. K. T., and DONOVAN, B., 1948, *Nature*, **162**, 336.
 EUCKEN, A., and FÖRSTER, F., 1934, *Nachr. Ges. Wiss. Gött. (Math. Phys. Klasse)*, **1**, 43.
 GRÜNEISEN, E., and GIELESSEN, J., 1936, *Ann. Phys., Lpz.*, **26**, 449; 1937, *Ibid.*, **28**, 225.
 DE HAAS, W. J., BLUM, J. W., and SCHUBNIKOW, L., 1935, *Physica, 's Grav.*, **2**, 907.
 HEAPS, C. W., 1926, *Phys. Rev.*, **27**, 764.
 JONES, H., 1936, *Proc. Roy. Soc. A*, **155**, 653.
 KAPITZA, P., 1928, *Proc. Roy. Soc. A*, **119**, 358.
 KAYE, G. W. C., 1939, *Proc. Roy. Soc. A*, **170**, 561.
 KOHLER, M., 1934, *Ann. Phys. Lpz.*, **20**, 891.
 MCLENNAN, J. C., ALLIN, E. J., and BURTON, A. C., 1932, *Phil. Mag.* [7], **14**, 508.
 PIETENPOL, W. B., and MILEY, H. A., 1929, *Phys. Rev.*, **34**, 1588.
 SCHUBNIKOW, L., and DE HAAS, W. J., 1930, *Commun. Phys. Lab. Leiden*, 207 and 210.
 SONDHEIMER, E. H., and WILSON, A. H., 1947, *Proc. Roy. Soc. A*, **190**, 435.
 STIERSTADT, O., 1933a, *Zeits. f. Physik*, **80**, 636.; 1933b, *Ibid.*, **85**, 310; 1933c, *Ibid.*, **85**, 697; 1934, *Ibid.*, **87**, 687; 1935a, *Ibid.*, **93**, 676; 1935b, *Ibid.*, **95**, 355.
 STRONG, J., 1940, *Modern Physical Laboratory Practice* (London: Blackie).
 THOMPSON, N., 1938, *Proc. Roy. Soc. A*, **164**, 24.

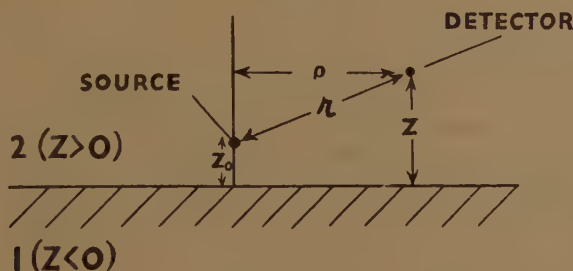
XXVII. Laplace Transform Solution of Two-medium Neutron Ageing Problem*.

By R. BELLMAN †, R. E. MARSHAK ‡, and G. M. WING §,
University of California, Los Alamos Scientific Laboratory,
Santa Fe, New Mexico ||.

[Received January 20, 1947.]

1. WE consider a point source of mono-energetic fast neutrons situated at a height z_0 (cf. fig. 1) above the interface between two semi-infinite plane media of different slowing down properties. We assume that age theory (Fermi 1936) holds for both media; thus, the equations for the slowing down densities, $\psi_1(\rho, z, u)$ and $\psi_2(\rho, z, u)$, are (it is convenient to use cylindrical coordinates):

Fig. 1.



$$\frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial \psi_1}{\partial \rho} \right) + \frac{\partial^2 \psi_1}{\partial z^2} = \frac{\partial \psi_1}{\partial t_1}, \quad \dots \quad (1.1.1)$$

$$\frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial \psi_2}{\partial \rho} \right) + \frac{\partial^2 \psi_2}{\partial z^2} = \frac{\partial \psi_2}{\partial t_2}, \quad \dots \quad (1.1.2)$$

where $t_{1,2} = \frac{\int_0^u l_{1,2}^2(u') du'}{3\xi_{1,2}[1 - \cos \Theta_{1,2}]}$ with $l(u)$ the mean free path for scattering,

* Parts of the solution first appeared in the declassified reports LADC—242 and LADC—243. Cf. R. E. Marshak, 1947, *Rev. Mod. Phys.* **19**, 185.

† Now in the Department of Mathematics, Stanford University.

‡ Now in the Department of Physics, The University of Rochester.

§ Now in the Department of Mathematics, Cornell University.

|| Communicated by R. E. Marshak.

ξ^* the average logarithmic energy loss, and $\overline{\cos \Theta}^*$ the average of the cosine of the angle of deflection in one collision. The variable u is equal to $\log (E_0/E)$, where E_0 is the energy of the source neutrons and E is the energy of interest.

Equations (1.1.1) and (1.1.2) are to be solved subject to the initial condition: point source

$$z=z_0 \geq 0, \quad \rho=0, \quad t_2=0 \quad . \quad . \quad . \quad . \quad . \quad (1.2)$$

and to the boundary conditions †:

$$\frac{l_1(u)}{\xi_1} \psi_1(\rho, z, u) = \frac{l_2(u)}{\xi_2} \psi_2(\rho, z, u) \quad \text{at } z=0 \quad . \quad . \quad (1.3.1)$$

(continuity of neutron density on plane $z=0$ for all ρ and u)

$$\frac{l_1^2(u)}{\xi_1[1-\overline{\cos \Theta}_1]} \frac{\partial \psi_1}{\partial z}(\rho, z, u) = \frac{l_2^2(u)}{\xi_2[1-\overline{\cos \Theta}_2]} \frac{\partial \psi_2}{\partial z}(\rho, z, u) \quad \text{at } z=0 \quad . \quad (1.3.2)$$

(continuity of normal neutron current on plane $z=0$ for all ρ and u).

If we now assume the scattering mean free paths in the two media to vary with u (*i. e.* energy) in the same way, we can write

$$D = \frac{l_2^2(u)}{\xi_2[1-\overline{\cos \Theta}_2]} \bigg/ \frac{l_1^2(u)}{\xi_1[1-\overline{\cos \Theta}_1]}, \quad . \quad . \quad . \quad (1.4.1)$$

where D is a constant greater than unity ‡. Also, we write

$$\alpha = \frac{\xi_1}{[1-\overline{\cos \Theta}_1]} \bigg/ \frac{\xi_2}{[1-\overline{\cos \Theta}_2]}. \quad . \quad . \quad . \quad (1.4.2)$$

It follows that $t_2 = Dt_1$; moreover, the boundary conditions (1.3.1) and (1.3.2) become

$$\psi_1(\rho, z, t_2) = \sqrt{D\alpha} \psi_2(\rho, z, t_2), \quad \left. \begin{array}{l} . \quad . \quad . \quad (1.5.1) \\ \text{at } z=0. \end{array} \right\} \quad . \quad . \quad . \quad (1.5.2)$$

$$\frac{\partial \psi_1}{\partial z}(\rho, z, t_2) = D \frac{\partial \psi_2}{\partial z}(\rho, z, t_2), \quad \left. \begin{array}{l} . \quad . \quad . \quad (1.5.1) \\ \text{at } z=0. \end{array} \right\} \quad . \quad . \quad . \quad (1.5.2)$$

Equations (1.1.1) and (1.1.2) together with the initial condition (1.2) and the boundary conditions (1.5.1) and (1.5.2) are formally equivalent to another physical problem, namely: given a point source releasing a pulse of heat in the presence of two semi-infinite substances of different densities and diffusivities§, what are the temperatures in the two media

* The quantities ξ and $\overline{\cos \Theta}$ are assumed to be constants independent of u ; this is strictly true for an element and approximately true for most mixtures.

† Equations (1.3.1) and (1.3.2) follow from the definition of the slowing down density ψ as the number of neutrons per unit volume per unit time reaching the age t .

‡ The subscripts 1 and 2 are chosen so that this is the case.

§ The diffusivity is the ratio of the thermal conductivity to the density.

as a function of time. This equivalence is true provided the following identification is made :

$$\begin{array}{ll} \psi_1 & \longrightarrow \text{temperature in medium 1,} \\ \sqrt{D\alpha} \psi_2 & \longrightarrow \text{temperature in medium 2,} \\ t_2/K_2 & \longrightarrow \text{time,} \\ D & \longrightarrow K_2/K_1, \\ \sqrt{D\alpha} & \longrightarrow \rho_1/\rho_2. \end{array}$$

In (1.6), K is the diffusivity of the medium (Carslaw), and ρ is the density.

2. The equations we must solve are, therefore ($t \equiv t_2$),

$$\left. \begin{array}{l} \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial \psi_1}{\partial \rho} \right) + \frac{\partial^2 \psi_1}{\partial z^2} = D \frac{\partial \psi_1}{\partial t} \quad (z < 0), \\ \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial \psi_2}{\partial \rho} \right) + \frac{\partial^2 \psi_2}{\partial z^2} = \frac{\partial \psi_2}{\partial t} \quad (z > 0), \end{array} \right\} \quad \dots \quad (2.1)$$

with the initial condition

(2.2.1) a point source at $z = z_0 \geq 0$, $\rho = 0$, $t = 0$, of unit strength, and the boundary conditions

$$\left. \begin{array}{l} \psi_1(\rho, z, t) = \sqrt{D\alpha} \psi_2(\rho, z, t), \\ \frac{\partial \psi_1}{\partial z}(\rho, z, t) = D \frac{\partial \psi_2}{\partial z}(\rho, z, t), \end{array} \right\} \text{at } z = 0. \quad \dots \quad (2.2.2)$$

Taking Laplace transforms with respect to t in order to eliminate the partial derivatives with respect to t , equations (2.1), (2.2.1) and (2.2.2) become

$$\left. \begin{array}{l} \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial \phi_1}{\partial \rho} \right) + \frac{\partial^2 \phi_1}{\partial z^2} = D \eta \phi_1 \quad (z < 0), \\ \frac{1}{\rho} \frac{\partial}{\partial \rho} \left(\rho \frac{\partial \phi_2}{\partial \rho} \right) + \frac{\partial^2 \phi_2}{\partial z^2} = \eta \phi_2 \quad (z > 0), \end{array} \right\} \quad \dots \quad (2.3)$$

(2.4.1) a point source at $z = z_0 \geq 0$, $\rho = 0$, of unit strength.

$$\left. \begin{array}{l} \phi_1(\rho, z, \eta) = \sqrt{D\alpha} \phi_2(\rho, z, \eta), \\ \frac{\partial \phi_1}{\partial z}(\rho, z, \eta) = D \frac{\partial \phi_2}{\partial z}(\rho, z, \eta), \end{array} \right\} \text{at } z = 0, \quad \dots \quad (2.4.2)$$

where

$$\begin{aligned} \phi_{1,2}(\rho, z, \eta) &= \int_0^\infty e^{-\eta t} \psi_{1,2}(\rho, z, t) dt, \\ &= \mathcal{L}_\eta \{ \psi_{1,2}(\rho, z, t) \}. \quad \dots \quad (2.5) \end{aligned}$$

If η is regarded as a parameter, then these equations are formally identical with those occurring in the Sommerfeld theory of the effect of a finitely conducting plane upon the radiation of an oscillating dipole.

The Sommerfeld method consists in taking solutions of the form (Stratton, a)

$$\left. \begin{aligned} \phi_1 &= \int_0^\infty f_1(\lambda) J_0(\lambda \rho) e^{z\sqrt{\lambda^2 + D\eta}} d\lambda & (z < 0), \\ \phi_2 &= \frac{e^{-r\sqrt{\eta}}}{4\pi r} + \int_0^\infty f_2(\lambda) J_0(\lambda \rho) e^{-z\sqrt{\lambda^2 + \eta}} d\lambda & (z > 0), \end{aligned} \right\} \quad (2.6)$$

where

$$r = \sqrt{\rho^2 + (z - z_0)^2}. \quad (2.7)$$

The representation (Stratton, b)

$$\frac{e^{-r\sqrt{\eta}}}{4\pi r} = \frac{1}{4\pi} \int_0^\infty \frac{J_0(\lambda \rho)}{\sqrt{\lambda^2 + \eta}} e^{-|z - z_0|\sqrt{\lambda^2 + \eta}} \lambda d\lambda \quad (2.8)$$

enables the expression for ϕ_2 to assume a more compact form

$$\phi_2 = \int_0^\infty J_0(\lambda \rho) \left\{ \frac{\lambda e^{-|z - z_0|\sqrt{\lambda^2 + \eta}}}{4\pi\sqrt{\lambda^2 + \eta}} + f_2(\lambda) e^{-z\sqrt{\lambda^2 + \eta}} \right\} d\lambda. \quad (2.9)$$

Using the boundary conditions (2.4.2)

$$\left. \begin{aligned} f_1(\lambda) &= \sqrt{D\alpha} \left[\frac{\lambda e^{-z_0\sqrt{\lambda^2 + \eta}}}{4\pi\sqrt{\lambda^2 + \eta}} + f_2(\lambda) \right], \\ f_2(\lambda) &= D \left[\frac{\lambda e^{-z_0\sqrt{\lambda^2 + \eta}}}{4\pi} - f_2(\lambda) \sqrt{\lambda^2 + \eta} \right]. \end{aligned} \right\} \quad (2.10)$$

Solving this pair of simultaneous equations

$$\left. \begin{aligned} f_1(\lambda) &= \frac{\sqrt{\alpha} D \lambda}{2\pi} \frac{e^{-z_0\sqrt{\lambda^2 + \eta}}}{[\sqrt{D}\sqrt{\lambda^2 + \eta} + \sqrt{\alpha}\sqrt{\lambda^2 + D\eta}]}, \\ f_2(\lambda) &= \frac{\lambda e^{-z_0\sqrt{\lambda^2 + \eta}}}{4\pi\sqrt{\lambda^2 + \eta}} \left[\frac{\sqrt{D}\sqrt{\lambda^2 + \eta} - \sqrt{\alpha}\sqrt{\lambda^2 + D\eta}}{\sqrt{D}\sqrt{\lambda^2 + \eta} + \sqrt{\alpha}\sqrt{\lambda^2 + D\eta}} \right], \end{aligned} \right\} \quad (2.11)$$

thus

$$\left. \begin{aligned} \phi_1 &= \frac{D\sqrt{\alpha}}{2\pi} \int_0^\infty J_0(\lambda \rho) \frac{e^{-z_0\sqrt{\lambda^2 + \eta} + z\sqrt{\lambda^2 + D\eta}}}{[\sqrt{D}\sqrt{\lambda^2 + \eta} + \sqrt{\alpha}\sqrt{\lambda^2 + D\eta}]} \lambda d\lambda, \\ \phi_2 &= \int_0^\infty J_0(\lambda \rho) \left\{ \frac{\lambda e^{-|z - z_0|\sqrt{\lambda^2 + \eta}}}{4\pi\sqrt{\lambda^2 + \eta}} \right. \\ &\quad \left. + \frac{\lambda e^{-(z + z_0)\sqrt{\lambda^2 + \eta}}}{4\pi\sqrt{\lambda^2 + \eta}} \left[\frac{\sqrt{D}\sqrt{\lambda^2 + \eta} - \sqrt{\alpha}\sqrt{\lambda^2 + D\eta}}{\sqrt{D}\sqrt{\lambda^2 + \eta} + \sqrt{\alpha}\sqrt{\lambda^2 + D\eta}} \right] \right\} d\lambda. \end{aligned} \right\} \quad (2.12)$$

For $z_0=0$, these solutions reduce to

$$\left. \begin{aligned} \phi_1 &= \frac{D\sqrt{\alpha}}{2\pi} \int_0^\infty \frac{J_0(\lambda\rho)e^{z\sqrt{\lambda^2+D\eta}}}{[\sqrt{D}\sqrt{\lambda^2+\eta}+\sqrt{\alpha}\sqrt{\lambda^2+D\eta}]} \lambda d\lambda, \\ \phi_2 &= \frac{\sqrt{D}}{2\pi} \int_0^\infty \frac{J_0(\lambda\rho)e^{-z\sqrt{\lambda^2+\eta}}}{[\sqrt{D}\sqrt{\lambda^2+\eta}+\sqrt{\alpha}\sqrt{\lambda^2+D\eta}]} \lambda d\lambda. \end{aligned} \right\} \quad (2.13)$$

Returning to ϕ_2 for $z_0>0$:

$$\begin{aligned} \phi_2 &= \int_0^\infty \frac{J_0(\lambda\rho)e^{-|z-z_0|\sqrt{\lambda^2+\eta}}}{4\pi\sqrt{\lambda^2+\eta}} \lambda d\lambda - \int_0^\infty \frac{J_0(\lambda\rho)e^{-(z+z_0)\sqrt{\lambda^2+\eta}}}{4\pi\sqrt{\lambda^2+\eta}} \lambda d\lambda, \\ &+ \frac{\sqrt{D}}{2\pi} \int_0^\infty \frac{J_0(\lambda\rho)e^{-(z+z_0)\sqrt{\lambda^2+\eta}}}{[\sqrt{D}\sqrt{\lambda^2+\eta}+\sqrt{\alpha}\sqrt{\lambda^2+D\eta}]} \lambda d\lambda \quad \dots \quad (2.14) \end{aligned}$$

as is seen by combining the second and third integrals.

To find ψ_1 and ψ_2 , it is necessary to find the Laplace inverse \mathcal{L}_η^{-1} of ϕ_1 and ϕ_2 . It is easy to find the inverses of the first two terms of (2.14). These two expressions represent intensities due to point sources of unit strength at $(0, z_0)$, $(0, -z_0)$ respectively, in an infinite medium. The difficulty lies in the third integral.

3. We shall give two methods for obtaining $\mathcal{L}_\eta^{-1}\{\phi_2\}$. The first method is capable of giving the inverse for all values of D, z, α . However, for the case of very large D ($D=\infty$ is actually treated, in a manner to be explained below), it yields quite complicated results. The second method is valid only for $D=\infty$ but yields much simpler results than the first. The first method will be used as far as practicable, and then the remaining cases will be treated by the second method.

In what follows, the continual interchange of the order of integration is easily justified by the absolute convergence of all integrals involved, provided we always take z as positive and consider the limiting case $z=0$, only after the Laplace inverse has been obtained.

The commutativity of the operator \mathcal{L}_η^{-1} with all integrations can be justified if we consider \mathcal{L}_η^{-1} as the integral operator

$$\frac{1}{2\pi i} \int_{b-i\infty}^{b+i\infty} e^{\eta v} \dots \dots \dots (3.2)$$

where b is any positive number. We can use this representation of \mathcal{L}_η^{-1} because the ϕ 's as functions of η are analytic functions of η for $\Re(\eta)>0$.

The third integral is the most troublesome, since, as mentioned above, the first two reduce to source terms, and their inverses can be readily found. Call the third integral B_2 .

It is easy to see that

$$B_2 = \frac{\sqrt{D}}{2\pi} \int_0^\infty J_0(\lambda\rho) \left[\int_0^\infty e^{-w[\sqrt{D}\sqrt{\lambda^2+\eta}+\sqrt{\alpha}\sqrt{\lambda^2+D\eta}]} - (z+z_0)\sqrt{\lambda^2+\eta} dv \right] \lambda d\lambda. \quad (3.3)$$

Thus

$$\mathcal{L}_\eta^{-1}\{B_2\} = \frac{\sqrt{D}}{2\pi} \int_0^\infty J_0(\lambda\rho) \times \left[\int_0^\infty \mathcal{L}_\eta^{-1}\{e^{-w[\sqrt{D}\sqrt{\lambda^2+\eta} + \sqrt{\alpha}\sqrt{\lambda^2+D\eta}] - (z+z_0)\sqrt{\lambda^2+\eta}}\}dw \right] \lambda d\lambda. \quad (3.4)$$

Consider the inverse transform

$$\mathcal{L}_\eta^{-1}\{e^{-\sqrt{\lambda^2+\eta}(w\sqrt{D}+z+z_0)} \cdot e^{-w\sqrt{\alpha}\sqrt{\lambda^2+D\eta}}\} = A(t, \lambda, w). \quad (3.5)$$

Now if

$$\mathcal{L}_\eta^{-1}\{f(\eta)\} = F(t); \quad \mathcal{L}_\eta^{-1}\{g(\eta)\} = G(t), \quad (3.6)$$

then the convolution theorem states that

$$\mathcal{L}_\eta^{-1}\{f(\eta)g(\eta)\} = \int_0^t F(u)G(t-u)du \quad (3.7)$$

and further

$$\mathcal{L}_\eta^{-1}\{f(\eta+a)\} = e^{-at}F(t). \quad (3.8)$$

The following inverse is well known (Doetsch),

$$\mathcal{L}_\eta^{-1}\{e^{-w\sqrt{\eta}}\} = \frac{we^{-w^3/4t}}{\sqrt{4\pi t^{3/2}}} \quad (3.9)$$

so that using (3.8)

$$\mathcal{L}_\eta^{-1}\{e^{-(w\sqrt{D}+z+z_0)\sqrt{\lambda^2+\eta}}\} = \frac{(w\sqrt{D}+z+z_0)e^{-\left[\frac{(w\sqrt{D}+z+z_0)^2}{4t} + \lambda^2 t\right]}}{\sqrt{4\pi} t^{3/2}}, \quad (3.10)$$

$$\mathcal{L}_\eta^{-1}\{e^{-w\sqrt{\alpha}\sqrt{\lambda^2+D\eta}}\} = \frac{w\sqrt{D\alpha}e^{-\left[\frac{D\alpha w^2}{4t} + \frac{\lambda^2 t}{D}\right]}}{\sqrt{4\pi} t^{3/2}}. \quad (3.11)$$

Therefore, using (3.7)

$$A(t, \lambda, w) = \frac{(w\sqrt{D}+z+z_0)w\sqrt{D\alpha}}{4\pi} \int_0^t \frac{e^{-\left[\frac{\lambda^2 u}{D} + \frac{w^2 D\alpha}{4u}\right]}}{u^{3/2}} \cdot \frac{e^{-\left[\lambda^2(t-u) + \frac{(w\sqrt{D}+z+z_0)^2}{4(t-u)}\right]}}{(t-u)^{3/2}} du. \quad (3.12)$$

Hence

$$\mathcal{L}_\eta^{-1}\{B_2\} = \frac{\sqrt{D}}{2\pi} \int_0^\infty \lambda J_0(\lambda\rho) \int_0^\infty A(t, \lambda, w)dw d\lambda. \quad (3.13)$$

Interchanging the order of integration

$$\mathcal{L}_\eta^{-1}\{B_2\} = \frac{D\sqrt{\alpha}}{8\pi^2} \int_0^\infty w(w\sqrt{D}+z+z_0) \times \left\{ \int_0^t \frac{e^{-\left[\frac{w^2 D\alpha}{4u} + \frac{(w\sqrt{D}+z+z_0)^2}{4(t-u)}\right]}}{u^{3/2}(t-u)^{3/2}} \cdot \left[\int_0^\infty \lambda J_0(\lambda\rho) e^{-\lambda^2 \left(\frac{u}{D} + t - u\right)} d\lambda \right] du \right\} dw. \quad (3.14)$$

Now (Watson),

$$\int_0^\infty \lambda J_0(\lambda \rho) e^{-t\lambda^2} d\lambda = \frac{e^{-\rho^2/4t}}{2t} \quad (3.15)$$

Thus

$$\int_0^\infty \lambda J_0(\lambda \rho) e^{-\lambda^2 \left[\frac{u}{D} + t - u \right]} d\lambda = \frac{e^{-\frac{\rho^2}{4 \left(\frac{u}{D} + t - u \right)}}}{2 \left(\frac{u}{D} + t - u \right)} \quad (3.16)$$

Applying (3.16), (3.14) reduces to

$$\begin{aligned} & \frac{D\sqrt{\alpha}}{16\pi^2} \int_0^\infty \frac{e^{-\frac{\rho^2}{4 \left(\frac{u}{D} + t - u \right)}}}{u^{3/2} (t-u)^{3/2} \left(\frac{u}{D} + t - u \right)} \\ & \times \left[\int_0^\infty w (w\sqrt{D} + z + z_0) e^{-\left[\frac{w^2 D \alpha}{4u} + \frac{(w\sqrt{D} + z + z_0)^2}{4(t-u)} \right]} dw \right] du. \end{aligned} \quad (3.17)$$

The next step is to evaluate the w integral. This is elementary, but tedious. Calling the bracketed expression $g_2(u, t, \rho, z)$:

$$\begin{aligned} g_2(t, u, \rho, z) &= \frac{2e^{-\frac{(z+z_0)^2 \alpha}{4(\alpha t - \alpha u + u)}}}{D(\alpha t - \alpha u + u)^{3/2}} \sqrt{u(t-u)} \\ & \times \left\{ e^{-\frac{(z+z_0)^2 u}{4(\alpha t - \alpha u + u)(t-u)}} \cdot \alpha(z+z_0) u^{1/2} (t-u)^{3/2} \right. \\ & + \frac{\sqrt{\pi}}{2} \left[1 - \operatorname{erf} \left\{ \frac{(z+z_0)\sqrt{u}}{2\sqrt{t-u}\sqrt{\alpha t - \alpha u + u}} \right\} \right] \\ & \times \left[u(t-u) \left(2 - \frac{(z+z_0)^2 \alpha}{\alpha t - \alpha u + u} \right) \right] \left. \right\}. \end{aligned} \quad (3.18)$$

Thus:

$$\mathcal{L}_\eta^{-1}\{B_2\} \equiv C_2(\rho, z, t) = \frac{D\sqrt{\alpha}}{16\pi^2} \int_0^t \frac{e^{-\frac{\rho^2}{4 \left(\frac{u}{D} + t - u \right)}}}{u^{3/2} (t-u)^{3/2} \left(\frac{u}{D} + t - u \right)} g_2(u, t, \rho, z) du. \quad (3.19)$$

The general solution for ψ_2 is then

$$\psi_2(\rho, z, t) = \frac{e^{-r_1^2/4t}}{(4\pi t)^{3/2}} - \frac{e^{-r_2^2/4t}}{(4\pi t)^{3/2}} + C_2(\rho, z, t), \quad (3.20)$$

where

$$\left. \begin{aligned} r_1^2 &= \rho^2 + (z - z_0)^2, \\ r_2^2 &= \rho^2 + (z - z_0)^2, \end{aligned} \right\}, \quad (3.21)$$

the first two terms having been obtained by applying (2.8) to (2.14) with the known inverse (3.9).

The same methods suffice for determining ψ_1 . The result is

$$\psi_1 = \frac{D^{3/2}\sqrt{\alpha}}{16\pi^2} \int_0^t \frac{q_1(u, t, \rho, z) e^{-\left[\frac{\rho^2}{4\left(\frac{u}{D} + t - u\right)} + \frac{1}{4u(t-u)}\right]}}{(t-u)^{3/2} u^{3/2} \left(\frac{u}{D} + t - u\right)} \times \left\{ z_0^2 u + D(t-u)z^2 - \frac{[uz_0 - z\sqrt{\alpha D(t-u)}]^2}{u + \alpha(t-u)} \right\} du, \quad (3.22)$$

where

$$q_1(u, t, \rho, z) = \frac{1}{A} \left\{ \frac{\sqrt{\pi}}{2} \left[\frac{\sqrt{\alpha D}}{A^2} \left(\frac{1}{2} + B^2 \right) - \frac{(z_0 \sqrt{\alpha} - z \sqrt{D})B}{A} - z z_0 \right] (1 - \operatorname{erf} B) + \frac{1}{2A} \left[z_0 \sqrt{\alpha} - z \sqrt{D} - \frac{B \sqrt{\alpha D}}{A} \right] e^{-B^2} \right\}, \quad (3.23)$$

$$A = \frac{\sqrt{D} \sqrt{u + \alpha(t-u)}}{4u(t-u)}, \quad B = \frac{uz_0 - z \sqrt{D \alpha(t-u)}}{4u(t-u) \sqrt{u + \alpha(t-u)}}. \quad (3.24)$$

(3.25) Reciprocity Relations. From the result stated in (3.20), it is clear that ψ_2 is a symmetric function of z and z_0 . This yields a reciprocity theorem which can be stated as follows :

Under the boundary conditions given, the intensity at a point (ρ_1, z_1) , $z_1 > 0$, due to a source at a point (ρ_2, z_2) , $z_2 > 0$, is equal to the intensity at the point (ρ_1, z_1) due to a source of the same strength at (ρ_2, z_2) .

4. The solution for ψ_2 , although quite complicated, can be computed numerically, using standard methods of numerical integration, since the integrand contains only algebraic and tabulated transcendental functions. In this section we consider various cases of physical importance, where certain simplifications are possible.

The work will be divided into five sections ; the first three will treat the case where the source is on the interface $z_0 = 0$, and the point of measurement is on the interface, $z = 0$; the remaining two will treat the general case. No simplification results from $z = 0$ if $z_0 \neq 0$.

$$z = z_0 = 0, \quad \alpha = 1, \quad 0 < D < \infty. \quad (4.1)$$

The result in (3.20) simplifies considerably, and

$$\psi_2 = \frac{1}{(4\pi t)^{3/2}} \int_0^t \frac{e^{-\frac{\rho^2}{4\left(\frac{u}{D} + t - u\right)}}}{\left(\frac{u}{D} + t - u\right)} du. \quad (4.1.1)$$

Changing variables, $w = \frac{u}{D} + t - u$, this becomes

$$\psi_2 = \left(\frac{D}{D-1} \right) \frac{1}{(4\pi t)^{3/2}} \left[\operatorname{Ei} \left(-\frac{\rho^2 D}{4t} \right) - \operatorname{Ei} \left(-\frac{\rho^2}{4t} \right) \right], \quad (4.1.2)$$

where

$$-\text{Ei}(-x) = \int_x^\infty \frac{e^{-y} dy}{y},$$

$$z = z_0 = 0, \quad \alpha = 1, \quad D = \infty. \quad . \quad . \quad . \quad . \quad . \quad (4.2)$$

In the following sections, the notation $D = \infty$ signifies that the principal term of the expression for large D has been taken. In this particular case, an error term can be obtained; that is, the next term in the asymptotic expansion can be found. Since, in the more general cases, it is quite difficult to obtain the error term, only the principal term will be given here, for the sake of consistency.

Taking the limit of (4.1.2) as $D \rightarrow \infty$

$$\psi_2 = -\frac{1}{(4\pi t)^{3/2}} \text{Ei}\left(-\frac{\rho^2}{4t}\right), \quad . \quad . \quad . \quad . \quad . \quad (4.2.1)$$

$$z = z_0 = 0, \quad 0 < \alpha < \infty (\alpha \neq 1), \quad D = \infty. \quad . \quad . \quad . \quad . \quad (4.3)$$

Letting $D \rightarrow \infty$ in (3.20)

$$\psi_2 = \frac{\sqrt{\alpha}}{(4\pi)^{3/2}} \int_0^t \frac{e^{-\frac{\rho^2}{4(t-u)}} du}{(\alpha t - \alpha u + u)^{3/2} (t-u)}. \quad . \quad . \quad . \quad . \quad . \quad (4.3.1)$$

A change of variable

$$t-u = \frac{t(1-v)}{1-\alpha}, \quad . \quad . \quad . \quad . \quad . \quad (4.3.2)$$

$$\psi_2 = \frac{\sqrt{\alpha}}{(4\pi t)^{3/2}} \int_\alpha^1 \frac{e^{\frac{\rho^2(\alpha-1)}{4t(1-v)}} dv}{v^{3/2}(1-v)}. \quad . \quad . \quad . \quad . \quad . \quad (4.3.3)$$

Taking the limit in (4.3.1) as $\alpha \rightarrow 1$, (4.2.1) is obtained.

$$0 < z + z_0 < \infty, \quad 0 < \alpha < \infty, \quad D = \infty. \quad . \quad . \quad . \quad . \quad (4.4)$$

Here another mode of procedure gives simpler results and seems interesting for its own sake.

It is quite difficult, due to the fact that we are working in transform space η , to evaluate the order of error in the inverse transform when approximations are made in the transform $\psi_{1,2}$. Therefore, we assume that to find the principal term in the asymptotic expansion of $\psi_{1,2}$ as $D \rightarrow \infty$, it is legitimate to find the principal term in the asymptotic expansion of $\phi_{1,2}$ as $D \rightarrow \infty$ and to take the inverse. As a check, it is to be noted that the results obtained in this way for the special case $z = z_0 = 0, \alpha = 1$, agree with the results obtained previously.

Clearly, we need consider only the expression for B_2 since the source terms in (2.14) are independent of D . For convenience, we set $z_1 = z + z_0$ in what follows.

From (2.14) *

$$\left. \begin{aligned} \lim_{D \rightarrow \infty} B_2 &= \frac{1}{2\pi} \int_0^\infty \frac{J_0(\lambda \rho) e^{-z_1 \sqrt{\lambda^2 + \eta}}}{[\sqrt{\lambda^2 + \eta} + \sqrt{\alpha \eta}] } \lambda d\lambda \\ &= \frac{1}{2\pi} \int_0^\infty \frac{J_0(\lambda \rho) e^{-z_1 \sqrt{\lambda^2 + \eta}}}{\sqrt{\lambda^2 + \eta} \left[1 + \frac{\sqrt{\alpha \eta}}{\sqrt{\lambda^2 + \eta}} \right]} \lambda d\lambda \\ &= \frac{1}{2\pi} \int_0^\infty \frac{J_0(\lambda \rho) e^{-z_1 \sqrt{\lambda^2 + \eta}}}{\sqrt{\lambda^2 + \eta}} \left[\sum_{k=0}^\infty \frac{(-1)^k (\sqrt{\eta \alpha})^k}{(\sqrt{\eta + \lambda^2})^k} \right] \lambda d\lambda, \end{aligned} \right\} \quad (4.4.1)$$

provided that for all λ

$$\left| \frac{\sqrt{\eta \alpha}}{\sqrt{\lambda^2 + \eta}} \right| < 1. \quad (4.4.2)$$

This will be true for all $|\alpha| < 1$, since η is non-negative. As above (cf. 2.8)

$$\int_0^\infty \frac{J_0(\lambda \rho) e^{-z_1 \sqrt{\lambda^2 + \eta}}}{\sqrt{\lambda^2 + \eta}} \lambda d\lambda = \frac{e^{-\sqrt{\eta} \sqrt{\rho^2 + z_1^2}}}{\sqrt{\rho^2 + z_1^2}}; \quad (\rho^2 + z_1^2) > 0. \quad (4.4.3)$$

To obtain higher powers of $\sqrt{\lambda^2 + \eta}$ in the denominator, we integrate (4.4.3) with respect to z_1 , over (z_1, ∞) . Thus

$$\int_0^\infty \frac{J_0(\lambda \rho) e^{-z_1 \sqrt{\lambda^2 + \eta}}}{(\sqrt{\lambda^2 + \eta})^{m+1}} \lambda d\lambda = \underbrace{\int_{z_1}^\infty dz_1 \dots \int_{z_1}^\infty}_{m \text{ times}} \frac{e^{-\sqrt{\eta} \sqrt{\rho^2 + z_1^2}}}{\sqrt{\rho^2 + z_1^2}} dz_1. \quad (4.4.4)$$

Letting

$$\zeta(z_1) = \frac{e^{-\sqrt{\eta} \sqrt{\rho^2 + z_1^2}}}{\sqrt{\rho^2 + z_1^2}}, \quad (4.4.5)$$

$$\begin{aligned} \int_0^\infty \frac{J_0(\lambda \rho) e^{-z_1 \sqrt{\lambda^2 + \eta}}}{[\sqrt{\lambda^2 + \eta} + \sqrt{\alpha \eta}] } \lambda d\lambda &= \zeta(z_1) - \sqrt{\eta \alpha} \int_{z_1}^\infty \zeta(z_1) dz_1 + (\sqrt{\eta \alpha})^2 \int_{z_1}^\infty dz_1 \int_{z_1}^\infty \zeta(z_1) dz_1 \\ &\quad + \dots + (-\sqrt{\eta \alpha})^m \underbrace{\int_{z_1}^\infty dz_1 \dots \int_{z_1}^\infty}_{m \text{ times}} \zeta(z_1) dz_1 + \dots \end{aligned} \quad (4.4.6)$$

provided that the series converges. This will certainly be so if

$$\sum_{m=0}^\infty |\sqrt{\eta \alpha}|^m \cdot \underbrace{\left| \int_{z_1}^\infty dz_1 \dots \int_{z_1}^\infty \zeta(z_1) dz_1 \right|}_{m \text{ times}} < \infty. \quad (4.4.7)$$

* Equations (4.4.1) can be obtained more directly by solving the second of equations (2.3) subject to the boundary condition.

But $(\rho^2 + z_1^2) > 0$, and the convergence follows since

$$|\sqrt{\eta\alpha}|^m \cdot \underbrace{\left| \int_{z_1}^{\infty} dz_1 \dots \int_{z_1}^{\infty} \zeta(z_1) dz \right|}_{m \text{ times}} < \frac{|\sqrt{\eta\alpha}|^m}{\max(\rho, z_1)} \underbrace{\int_{z_1}^{\infty} dz_1 \dots \int_{z_1}^{\infty} e^{-z_1 \sqrt{\eta}} dz_1}_{m \text{ times}} \\ < \frac{|\sqrt{\alpha}|^m}{\max(\rho, z_1)} \dots \dots \dots (4.4.8)$$

Now consider :

$$\int_{z_1}^{\infty} e^{-z_1 \sqrt{\eta\alpha}} \zeta(z_1) dz_1 = e^{-z_1 \sqrt{\eta\alpha}} \int_{z_1}^{\infty} \zeta(z_1) dz_1 - \sqrt{\eta\alpha} e^{-z_1 \sqrt{\eta\alpha}} \int_{z_1}^{\infty} dz_1 \int_{z_1}^{\infty} \zeta(z_1) dz_1 \\ + \dots + (-\sqrt{\eta\alpha})^m e^{-z_1 \sqrt{\eta\alpha}} \underbrace{\int_{z_1}^{\infty} dz_1 \dots \int_{z_1}^{\infty} \zeta(z_1) dz_1}_{m \text{ times}} + \dots \quad (4.4.9)$$

which is obtained by repeated integrations by parts. The right-hand side converges by (4.4.7). Comparing (4.4.6 and (4.4.9), we get :

$$\frac{1}{2\pi} \int_0^{\infty} \frac{J_0(\lambda\rho) e^{-z_1 \sqrt{\lambda^2 + \eta}}}{[\sqrt{\lambda^2 + \eta} + \sqrt{\eta\alpha}]} \lambda d\lambda = \frac{\zeta(z_1)}{2\pi} - \frac{e^{z_1 \sqrt{\eta\alpha}} \sqrt{\eta\alpha}}{2\pi} \int_{z_1}^{\infty} \zeta(z_1) e^{-z_1 \sqrt{\eta\alpha}} dz_1. \quad (4.4.10)$$

This identity has been established under the restriction $|\alpha| < 1$. For the values of λ and η that arise, both sides, considered as functions of $\sqrt{\alpha}$, are analytic functions of $\sqrt{\alpha}$ for $\Re(\sqrt{\alpha}) > 0$. Therefore, by the principle of analytic continuation, the identity proved only for $|\alpha| < 1$ holds for all positive α .

The inverse, \mathcal{L}_{η}^{-1} , can be obtained in a straight-forward manner, and thus

$$\lim_{D \rightarrow \infty} C_2 = \frac{1}{4(\pi t)^{3/2}} \left[e^{-\frac{(\rho^2 + z_1^2)}{4t}} + \sqrt{\alpha} \int_{z_1}^{\infty} \frac{\frac{\partial}{\partial f} (f e^{-\frac{f^2}{4t}})}{\sqrt{\rho^2 + w^2}} dw \right] \quad (4.4.11)$$

where

$$f = \sqrt{\alpha}(w - z_1) + \sqrt{\rho^2 + w^2}. \quad (4.4.12)$$

Explicitly

$$\lim_{D \rightarrow \infty} C_2 = \frac{1}{4(\pi t)^{3/2}} \left[e^{-\frac{(\rho^2 + z_1^2)}{4t}} + \sqrt{\alpha} \int_{z_1}^{\infty} \frac{1}{\sqrt{\rho^2 + w^2}} \right. \\ \left. \times \left\{ 1 - \frac{[\sqrt{\alpha}(w - z_1) + \sqrt{\rho^2 + w^2}]^2}{2t} \right\} \cdot e^{-\frac{[\sqrt{\alpha}(w - z_1) + \sqrt{\rho^2 + w^2}]^2}{4t}} dw \right]. \quad (4.4.13)$$

Performing a change of variable

$$u = \sqrt{\rho^2 + w^2} + \sqrt{\alpha}(w - z_1), \quad (4.4.14)$$

$$\lim_{D \rightarrow \infty} C_2 = \frac{1}{4(\pi t)^{3/2}} \left[e^{-\frac{(\rho^2 + z_1^2)}{4t}} + \sqrt{\alpha} \int_{\sqrt{\rho^2 + z_1^2}}^{\infty} \frac{e^{-\frac{u^2}{4t}} \left(1 - \frac{u^2}{2t}\right) du}{[u^2 + \alpha z_1^2 + 2z_1 u \sqrt{\alpha} + \rho^2(\alpha - 1)]^{1/2}} \right] \quad (4.4.15)$$

$$0 < z + z_0 < \infty, \quad D = \infty, \quad \alpha = 1. \quad (4.5)$$

Here we have

$$\lim_{D \rightarrow \infty} C_2 = \frac{1}{4(\pi t)^{3/2}} \left[e^{-\frac{(\rho^2 + z_1^2)}{4t}} + \int_{\sqrt{\rho^2 + z_1^2}}^{\infty} \frac{e^{-\frac{u^2}{4t}} \left(1 - \frac{u^2}{2t}\right) du}{u + z_1} \right] \quad (4.5.1)$$

This can be simplified slightly to

$$\lim_{D \rightarrow \infty} C_2 = \frac{1}{4(\pi t)^{3/2}} \left[\frac{z_1 \sqrt{\pi}}{2\sqrt{t}} \left\{ 1 - \operatorname{erf} \left(\sqrt{\frac{\rho^2 + z_1^2}{4t}} \right) \right\} + \frac{(2t - z_1^2)}{2t} \int_{[z_1 + \sqrt{\rho^2 + z_1^2}]}^{\infty} \frac{e^{-\frac{(u - z_1)^2}{4t}}}{u} du \right] \quad (4.5.2)$$

For $z_1 = 0$, (4.5.2) reduces to

$$\lim_{D \rightarrow \infty} C_2 = \frac{1}{(4\pi t)^{3/2}} \operatorname{Ei} \left(-\frac{\rho^2}{4t} \right), \quad (4.5.3)$$

which agrees with the previous result (4.2.1).

Finally, it should be noted that $t^{3/2}\psi_{1,2}$ are in all cases functions only of $\left(\frac{\rho^2}{4t}\right)$ and $\left(\frac{z^2}{4t}\right)$. Computation can be simplified considerably by taking advantage of this fact.

This work was carried out under contract between the University of California and the Manhattan District, Corps of Engineers, War Department, U.S.A.

REFERENCES.

- CARSLAW, H. S., *Conduction of Heat in Solids*, p. 8.
 DOETSCH, G., *Theorie und Anwendung der Laplace-Transformation*, p. 402.
 FERMI, E., 1936, *Ric. Scient.* (2) **7**, 13.
 STRATTON, J. A., a, *Electromagnetic Theory*, p. 573; b, p. 576.
 WATSON, G. N., *A Treatise on the Theory of Bessel Functions*, p. 393.

XXVIII. *Note on the Effect of Impurities and Cold Work on the Thermoelectric Power of Aluminium.*

By J. K. GALT*,

H. H. Wills Physical Laboratory, University of Bristol †.

[Received September 23, 1948.]

ABSTRACT.

Recent observations of Crussard on the thermoelectric power of impure aluminium are discussed theoretically. Dissolved impurities produce a change in the thermoelectric power of pure Al which can be understood in terms of the electronic band structure of Al. When the impure Al is cold worked, the variation of its thermoelectric power with time and temperature is changed. An explanation of this in terms of the theory of dislocations is proposed.

1. INTRODUCTION.

CRUSSARD (1948, 1949) has recently observed two interesting facts about the thermoelectric power of aluminium. The first is that its thermoelectric power is increased by dissolving small concentrations of a metal from a column preceding that of aluminium in the periodic table. Conversely the thermoelectric power is decreased by dissolving other metals in Al. In particular among the metals in columns following Al in the periodic table Crussard observes a change varying from -1.5 per cent for Si to -27 per cent for Mn for a dissolved impurity content of 0.1 per cent. The second interesting observation is that if an impurity is dissolved in a metal, and the system is then quenched and cold worked, the variation of the thermoelectric power S as the temperature is slowly raised again is given by a curve like that labelled "cold worked metal" in fig. 1. If, on the other hand, the metal is given the same treatment but is not cold worked before the temperature is raised again the variation is given by a curve like that labelled "metal not cold worked" in fig. 1. The axis of abscissæ in fig. 1 corresponds to the thermoelectric power of the pure metal.

It is our purpose in this note to present theoretical interpretations of these observations.

* Now at the Bell Telephone Laboratories, New York.

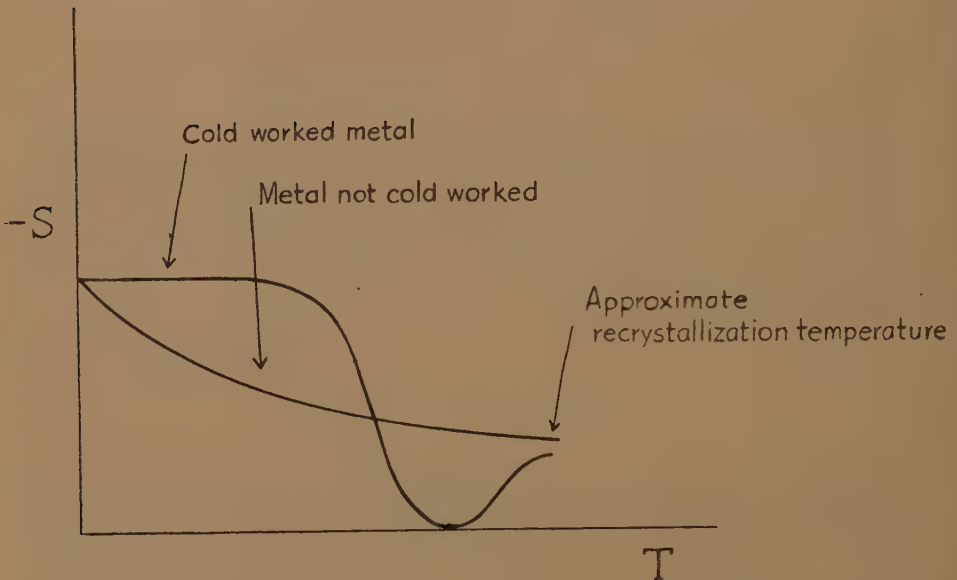
† Communicated by Prof. N. F. Mott, F.R.S.

2. EFFECT OF DISSOLVED IMPURITIES.

In the first place, we wish to point out that the effect of dissolved impurities may be explained qualitatively on the basis of the band structure of Al as described by Matyas (1948) in the same way that Mott (1936 a, b) has used the structure of the *s* and *d* bands in the transition metals to explain their thermoelectric power.

According to Matyas the important features of the band structure of Al are : (1) The three valence electrons which are in *s* and *p* states in the free atom form two bands with overlapping energies. (2) The Fermi

Fig. 1.



energy is near the top of the lower band, so that this band (the *s* band) is almost, but not quite, full. (3) The density of states near the top of the *s* band is large. We may illustrate this state of affairs by the diagram in fig 2, where $N(E)$ is the number of states between E and $E+dE$, ζ is the Fermi energy and E_0 is the highest energy in the *s* band. The similarity between this structure and that of the *s* and *d* bands in the transition metals is obvious.

The thermoelectric power is given by (Mott and Jones 1946 b) :—

$$S = \frac{\pi^2 k^2 T}{3e} \frac{\partial}{\partial E} [\log \sigma(E)]_{E=\zeta}, \quad (1)$$

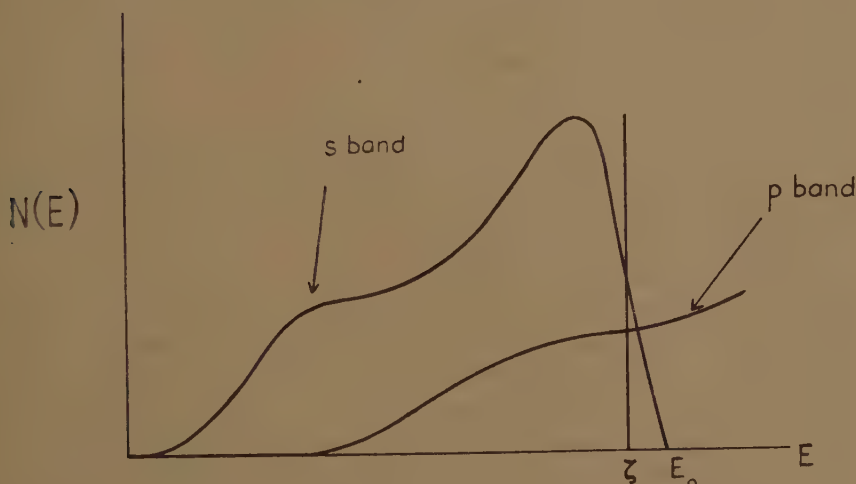
where $\sigma(E)$ is the conductivity when the maximum energy of the electrons is E . Now if we write σ as the sum of contributions from the *s* and *p* bands, σ_s and σ_p , and use a prime to indicate differentiation with respect to E , we see that S is proportional to $(\sigma'_s + \sigma'_p)/(\sigma_s + \sigma_p)$. But $\sigma_p \gg \sigma_s$ since the

number of free electrons in the p band is much larger than the number of holes in the s band. We therefore write :

$$S = \frac{\pi^2 k^2 T}{3e} \left(\frac{\sigma'_p}{\sigma_p} + \frac{\sigma'_s}{\sigma_s} \right)_{E=\zeta} = S_1 + S_2. \quad (2)$$

To calculate S_1 we note that $\sigma_p = n_p e^2 \tau_p / \mu_p$, where n_p is the effective number of free electrons per atom in the p band, e is the electronic charge, μ_p is the effective mass of an electron in the p band (almost equal to the mass of a free electron) and τ_p is the relaxation time as usually defined in conductivity theory. Now we neglect the variation of μ_p in differentiating with respect to E , and we obtain $\sigma'_p / \sigma_p = \tau'_p / \tau_p + n'_p / n_p$. For a band structure essentially

Fig. 2.



like that illustrated in fig. 2, Mott has derived the relation (Mott 1936 b) :

$$\frac{1}{\tau_p} = \text{const.} [A N_p(E) + B N_s(E)], \quad (3)$$

where A and B are constants which depend on the wave functions in the two bands. The effect of transitions from the s to the p band on conductivity has been neglected here, but the second term in (3) arises out of transitions from the p to the s band. We are encouraged to make this approximation because of the high conductivity of Al. For this same reason we expect the second term in (3) to be somewhat smaller than the first, but not negligible.

In deriving (3), Mott has also assumed that the energy surfaces are spheres, and we may therefore write $N_p(E) \sim \mu_p^{\frac{3}{2}} E^{\frac{1}{2}}$ and $N_s(E) \sim \mu_s^{\frac{3}{2}} (E_0 - E)^{\frac{1}{2}}$ where μ_s and μ_p are the electron masses in the s and p bands respectively.

Replacing σ'_p/σ_p by $\tau'_p/\tau_p + n'_p/n_p$ we now obtain for S_1 :

$$S_1 = \frac{\pi^2 k^2 T}{6e\zeta} \left[\frac{1 + \left(\frac{\mu_s}{\mu_p}\right)^{\frac{1}{2}} \frac{B}{A} \left(\frac{\zeta}{E_0 - \zeta}\right)^{\frac{1}{2}}}{1 + \left(\frac{\mu_s}{\mu_p}\right)^{\frac{1}{2}} \frac{B}{A} \left(\frac{E_0 - \zeta}{\zeta}\right)^{\frac{1}{2}}} + 3 \right], \quad \dots \quad (4)$$

since the electrons in the p band are almost free and $n_p \sim E^{\frac{1}{2}}$.

It is difficult to calculate S_2 in this way, as we do not know how τ_s and τ_p vary with E , but it is clear that S_2 is not large compared to S_1 : We can easily show (Mott and Jones 1936 a) that $S \sim -(\zeta - E_0)^x$, where x is determined by the dependence of τ_s on E . From this we see that S_2 may be either positive or negative. If it is positive it is smaller than S_1 , since the thermoelectric power of Al is negative, as is S_1 . If S_2 is negative, on the other hand, it cannot be appreciably larger than S_1 or the thermoelectric power of Al would be large compared to that of metals with only one unfilled band. This is obvious when it is remembered that S_1 includes the mechanism which gives rise to the whole thermoelectric power of Na, for example.

The fact that $S_2 \sim -(\zeta - E_0)^x$ also tells us that changes in ζ produced by dissolving impurities in Al lead to relatively small changes in S_2 when $(E_0 - \zeta)$ is small, as it is here. We therefore treat S_2 as a constant somewhat smaller than S_1 when we are considering the change in S due to small concentrations of dissolved impurities.

In order to explain Crussard's observations, therefore, we examine the variation of S_1 with ζ as given by (4). Since the second term in (3) is the smaller, the second term in the denominator of the first term in (4) is probably less than 1. But the Matyas analysis (see fig. 2) shows that the Fermi surface is very near E_0 , and hence $(E_0 - \zeta) < \zeta$. Consequently the second term in the numerator of the first term in (4) is probably larger than 1, and hence small changes in ζ may produce substantial changes in S because of the $[\zeta/(E_0 - \zeta)]^{\frac{1}{2}}$ factor. The thermoelectric power of pure Al and Matyas' estimate of ζ are also consistent with this conclusion. As a percentage, the changes are even larger if S_2 is positive. Furthermore the sign of the variation obtained by Crussard is predicted by (4), whatever the relative magnitudes of the various terms, for if we dissolve a metal from a column following Al in the periodic table, ζ increases and S becomes more negative (note that e is negative). If a metal from one of the preceding columns is dissolved ζ decreases and (4) predicts an increase in S . Thus, although the exact form of the variation of S with ζ cannot be specified, both the sign and the magnitude of the observed variations with small concentrations of impurities seem reasonable in terms of (4). Furthermore, if we dissolve small equal concentrations, we expect ζ to vary nearly linearly as we dissolve metals from succeeding columns in a row of the periodic table. If this is true, Crussard's results indicate that the variation of S with ζ becomes more rapid as ζ increases, as we should expect from (4).

3. EFFECT OF COLD WORK.

We interpret the observations illustrated in fig. 1 in terms of the theory of dislocations. In the case of the sample which is not cold worked it seems reasonable, as Crussard and his group have pointed out (Crussard 1949) that the variation with temperature (as well as a slow variation with time) is due to the diffusion of solute atoms into the few dislocations which are present even without cold work. From a picture of a cross-section of the usual model of a dislocation (Taylor 1934) it can be seen that solute atoms, because they are slightly different in size from the solvent atoms, tend to diffuse into one side or the other of a dislocation in order to relieve the stresses present in the dislocation. Which side of the dislocation the solute atoms diffuse toward will, of course, depend on whether they are larger or smaller than the solvent atoms, but this is not important for our purposes. Since, however, the two types of atoms are not greatly different in size several solute atoms are required in each atomic plane of the dislocation in order to relieve the stress, and the solute therefore precipitates out into a cylinder along the dislocation and the electrons associated with the solute atoms are removed from the Fermi sea of the metal as a whole. As a result, the thermoelectric power tends toward that of the pure solvent metal, as shown in fig. 1.

When the sample is cold worked before observations are made on it, however, we see from fig. 1 that the thermoelectric power behaves quite differently as the temperature is raised. The following interpretation of this behaviour is proposed. We note first that time as well as temperature is a variable in the observations. Now when the metal is first quenched and cold worked, the dislocations are large in number and randomly arranged, so that even if the solute atoms diffuse into dislocations they are still more or less randomly distributed in the solvent metal and there are not very many in any one dislocation. Therefore at first the electrons associated with the solute atoms remain in the Fermi sea of the metal as a whole, and the thermoelectric power remains constant. As time passes and the temperature is raised, however, recovery occurs in the solvent metal. As Mott (1948) has pointed out, this means that the dislocations arrange themselves into walls. When this happens, of course, the solute atoms form a three dimensional precipitate in the neighbourhood of these walls as they diffuse into dislocations, and again the thermoelectric power approaches that of the pure metal. We thus associate the flat part of the curve with the time delay found by Anderson and Mehl (1945) in the recovery process, and the rather sharp drop off with the occurrence of recovery. As the temperature is raised still higher, the solute dissolves again of course, especially at the recrystallization temperature where we might expect some dislocations to disappear, and this makes the thermoelectric power again deviate from that for the pure metal.

The author wishes to thank Professor N. F. Mott for suggesting this problem and Professor Mott and Dr. A. B. Bhatia for enlightening

discussions on it. He also wishes to express his gratitude to the U.S. National Research Council for a fellowship which has enabled him to spend a year in Bristol.

REFERENCES.

- ANDERSON, W. A., and MEHL, R. F., 1945, *Trans. Am. Inst. Mining and Metallurgical Eng. (Inst. of Metals Div.)*, **161**, 140.
 CRUSSARD, C., 1948, *Report of a Conference on the Strength of Solids* (held at H. H. Wills Physical Laboratory 7th-9th July, 1947) (London: Phys. Soc.), p. 119; 1949, Private Communication and a paper soon to appear in *Revue de Metallurgie*.
 MATYAS, Z., 1948, *Phil. Mag.* [7], **39**, 429.
 MOTT, N. F., 1936 a, *Proc. Roy. Soc. A*, **153**, 699; 1936 b, *Ibid.*, **156**, 368; 1948, *Report on Int. Congress on Physics of Metals in Amsterdam*, July; soon to appear in *Physica*. CAHN, R. W., 1949, a paper soon to appear in *Proc. Phys. Soc.*
 MOTT, N. F., and JONES, H., 1936 a, *Theory of the Properties of Metals and Alloys* (Oxford: University Press), p. 263, eq. (55); 1936 b, *Ibid.*, p. 310.
 TAYLOR, G. I., 1934, *Proc. Roy. Soc. A*, **145**, 362.

XXIX. *Distribution of Reaction Times for Turbulent Flow in Cylindrical Reactors.*

By R. C. L. BOSWORTH, Ph.D., D.Sc.*

[Received April 24, 1947.]

1. INTRODUCTION.

WITH respect to the time available for the progress of any chemical reaction the passage of a reaction mixture through a cylindrical vessel can be specified not by a single reaction time, but by a population of reaction times exhibiting a certain frequency distribution curve. In a recent paper the author (Bosworth 1947) derived the form of this distribution curve for the special case of laminar flow and showed how this could be modified by diffusion. The present paper will take up the case of turbulent flow, also through a cylindrical reactor.

The deduction in the case of laminar flow took as a starting point the Poiseuille equation for the parabolic distribution of velocities across a pipe. In the case of turbulent flow, while there is available no corresponding expression formulated from fundamental principles, there are a number of empirical equations for the distribution of velocities across a pipe which appear to hold with a tolerable degree of precision. Thus Prandtl (Ewald, Pöschl, and Prandtl 1930) states that "*—if Reynolds' number is not too large the velocity of flow along a smooth boundary is proportional to the seventh root of the distance from the boundary. For larger values of the Reynolds number it has been observed to be proportional to*

* Communicated by the Author.

the eighth root. Lower roots, e. g., the fifth root are obtained in the case of rough boundaries." The seventh root law of distribution of velocities has been used by Sutton (1934), who in particular showed that this leads to the Blasius form of the equation for the pressure drop in a pipe (pressure drop proportional to the seven-quarters power of the velocity) which for smooth pipes is valid over a range of Reynolds' numbers from 2000 to 100,000. Von Karman (1934) developed a theory of the turbulent flow pattern which gives the velocity v distance r from the axis of a pipe of radius R in terms of the maximum velocity v_0 in the pipe and the so-called friction velocity u^* , viz.,

$$\frac{v_0 - v}{u^*} = \frac{1}{0.38} \left\{ \ln(1 - \sqrt{r/R}) - \sqrt{r/R} \right\}. \quad (1)$$

Prandtl (1933) has suggested the simpler form

$$\frac{v_0 - v}{u^*} = 2.5 \ln \frac{R}{R-r} = 5.75 \log \frac{R}{R-r}. \quad (2)$$

Both these expressions make $v=v_0$ when $r=0$, but also made $v \rightarrow -\infty$ when $r \rightarrow R$, whereas physically $v=0$ when $r=R$ unless there is a definite slip at the walls. In other words expressions (1) and (2) clearly do not apply to the distribution of velocities across the laminar film near the pipe walls.

For the purposes of the present problem the power law empirical relationships are more convenient. Sutton (1934) for example took for the distribution law,

$$v = v_0 \left(\frac{R-r}{R} \right)^{1/7}, \quad (3)$$

which holds with tolerable accuracy over a considerable range of Reynolds' numbers. (The convention that r can only be positive must, however, rigorously be maintained.) In order to make the treatment slightly more general we will take as the distribution law

$$v = v_0 \left(\frac{R-r}{R} \right)^{1/n}, \quad (4)$$

where n is a function of the Reynolds number (Re), and varies also with the relative roughness of the walls of the vessel concerned. For relatively smooth pipes

$$\begin{aligned} n &= 7, & 2000 \leq Re \leq 100,000, \\ n &= 8, & Re > 100,000. \end{aligned}$$

2. DISTRIBUTION OF "STATISTICAL STREAMLINE" POINTS.

Diffusive mixing of a fluid flowing turbulently proceeds considerably faster than in the same fluid under laminar flow, and consequently calculation of the distribution function for the different molecules with respect to their times of transit through a given length of pipe, without

taking diffusion into account, is scarcely justified. However, as in the earlier paper, we will proceed by first calculating the distribution function for reaction times of imaginary points fixed relative to the statistical streamlines and then consider later how actual molecules will move under eddy diffusion with respect to these statistical streamline points.

The velocity distribution across the pipe is to be taken as

$$v = v_0 \left(\frac{R-r}{R} \right)^{1/n} \quad \dots \quad (4)$$

If L is the length of the pipe, the time τ taken by the point distance r from the axis to pass through the pipe is given by

$$\tau = \frac{L}{v} = \frac{L}{v_0} \left(\frac{R}{R-r} \right)^{1/n}, \quad \dots \quad (5)$$

which on differentiation becomes

$$\frac{d\tau}{dr} = \frac{L}{n v_0} \left(\frac{R}{R-r} \right)^{1/n} \frac{1}{R-r} \quad \dots \quad (6)$$

The volume (dQ) of fluid which in unit time enters a hollow cylinder of internal radius r and external radius $r+dr$, is given by

$$dQ = 2\pi r v dr, \quad \dots \quad (7)$$

this volume is characterized by the reaction time τ of equation (5). The total volume Q entering the pipe in unit time is

$$\begin{aligned} Q &= \int_0^R 2\pi r v dr \\ &= \frac{2\pi v_0}{R^{1/n}} \int_0^R r (R-r)^{1/n} dr \\ &= \frac{2\pi n^2 v_0 R^2}{(n+1)(2n+1)} \quad \dots \quad (8) \end{aligned}$$

The fraction $F_r dr$ of the total molecular stream which enters the hollow cylinder r to $r+dr$ and is characterized by the reaction time τ is given by

$$\begin{aligned} F_r dr &= 2\pi r v dr / Q \\ &= \frac{(n+1)(2n+1)}{n^2} \frac{r}{R^2} \left(\frac{R-r}{R} \right)^{1/n} dr \quad \dots \quad (9) \end{aligned}$$

The fraction $F_\tau d\tau$ of the stream exhibiting a reaction time lying between τ and $\tau+d\tau$ is thus

$$\begin{aligned} F_\tau d\tau &= F_r \frac{1}{d\tau/dr} d\tau \\ &= \frac{(n+1)(2n+1)}{n} \frac{v_0}{L} \left(\frac{R-r}{R} \right)^{2/n} \frac{r(R-r)}{R^2} d\tau, \end{aligned}$$

which on substitution in equation (5) becomes

$$F_{\tau} = \frac{(n+1)(2n+1)}{n} \tau^{-n-2} \left(\frac{L}{v_0} \right)^{n+1} \left(1 - \frac{L^n}{v_0^n \tau} \right).$$

Now $\frac{L}{v_0} = \tau_0$, the minimum reaction time or time for the passage of a point on the central core of fluid. Since no point travels faster than one on the central core we have

$$F_{\tau} = 0 \quad \text{for all } \tau < \tau_0.$$

Consequently the final expression for F_{τ} reads

$$F_{\tau} d\tau = H(\tau - \tau_0) \frac{(n+1)(2n+1)}{n} \frac{\tau_0^{n+1}}{\tau^{n+2}} \left(1 - \frac{\tau_0^n}{\tau^n} \right) d\tau. \quad (10)$$

If we adopt the notation used before and write τ^* for the relative reaction time, *i. e.*,

$$\tau^* = \tau / \tau_0,$$

we get

$$F_{\tau^*} = H(\tau^* - 1) \frac{(n+1)(2n+1)}{n} \tau^{*-n-2} (1 - \tau^{*-n}). \quad (11)$$

Thus when $n=7$

$$F_{\tau^*} = H(\tau^* - 1) \frac{120}{7} \tau^{*-9} (1 - \tau^{*-7}). \quad (11a)$$

when $n=8$

$$F_{\tau^*} = H(\tau^* - 1) \frac{153}{8} \tau^{*-10} (1 - \tau^{*-8}), \quad (11b)$$

and when $n=5$

$$F_{\tau^*} = H(\tau^* - 1) \frac{66}{5} \tau^{*-7} (1 - \tau^{*-5}). \quad (11c)$$

The mean value for the relative reaction time τ_m^* is given by

$$\begin{aligned} \tau_m^* &= \int_1^{\infty} \tau^* F_{\tau^*} d\tau^* \\ &= \frac{(n+1)(2n+1)}{2n^2} \end{aligned} \quad (12)$$

for

$$n=7 \quad \tau_m^* = 1.22,$$

$$n=8 \quad \tau_m^* = 1.19,$$

$$n=5 \quad \tau_m^* = 1.32.$$

For streamline flow it was shown in a previous paper that

$$\tau_m^* = 2.0,$$

so that, as might have been expected, the mean reaction time under conditions of turbulent flow is much more nearly equal to the minimum time than in streamline flow. In other words the spread of the population of reaction times is less under turbulent conditions, even when the effects of turbulent diffusion are neglected.

3. MAGNITUDE OF THE COEFFICIENT OF TURBULENT DIFFUSION.

Diffusion under turbulent conditions differs from diffusion under conditions of laminar flow in that the effective diffusion coefficient varies with the distance from the wall ($R-r$). The mixing length l , or the average distance an eddy travels before absorption in the rest of the fluid, is practically constant over the centre core of the fluid, but near the walls becomes proportional to the distance from the wall.

The average component of the eddy motion perpendicular to the wall (\bar{v}_y) also varies with the distance y from the walls and, according to Prandtl (1925), we may write

$$\bar{v}_y' = l \frac{d\bar{v}}{dy}, \quad \dots \dots \dots (13)$$

where \bar{v} is the average velocity parallel to the axis of the pipe. The extended Reynolds' analogy states that, *for a medium in turbulent motion the ratio of the momentum lost per unit time by skin friction to the total momentum losable is the same as the ratio of the amount of matter removed by the fluid from the wall to the amount which would have been removed had the fluid been allowed to come into equilibrium with the walls.* In other words the turbulent diffusivity D_t may be equated to $l\bar{v}_y'$ or

$$D_t = l^2 \frac{d\bar{v}}{dy}.$$

As an expression for the mixing length (l) von Karman (1934) has given

$$l = 0.4 \frac{(dv/dy)}{(d^2v/dy^2)}, \quad \dots \dots \dots (14)$$

so that we have for the coefficient of turbulent diffusion

$$D_t = 0.16 \frac{(dv/dy)^3}{(d^2v/dy^2)^2} \cdot \dots \dots \dots (15)$$

The values of dv/dy and d^2v/dy^2 obtained from equation (4) read

$$\frac{dv}{dy} = \frac{v_0}{R^{1/n}} \frac{1}{n} y^{\overline{1-n/n}} \quad \dots \dots \dots (4a)$$

and

$$\frac{d^2v}{dy^2} = - \frac{v_0}{R^{1/n}} \frac{1}{n} \cdot \frac{n-1}{n} \cdot y^{\overline{1-2n/n}}, \quad \dots \dots \dots (4b)$$

which on substitution in equation (15) gives

$$D_t = \frac{0.16 n}{(n-1)^2} \left(\frac{y}{R} \right)^{1/n} v_0 y \quad (16)$$

as an expression for the coefficient of turbulent diffusion as a function of the distance y from the walls.

The product $D_t \tau$, with τ as given by equation (5) occurs frequently in all expressions dealing with the modifications produced by diffusion in a reactor. The magnitude of the quantity is given by

$$D_t \tau = \frac{0.16 n}{(n-1)^2} L y. \quad (17)$$

This expression may be rewritten as

$$D_t \tau = \frac{R-r}{2b^2 R}, \quad (18)$$

where $R-r$ is written for y and b is defined by

$$b^2 = \frac{(n-1)^2}{0.32 n L R}, \quad (19)$$

and is thus a quantity of physical dimensions $1/R$. Since L , the length of the reaction zone is usually at least $20R$ (Bosworth 1948) it follows that the magnitude of bR is somewhat less than unity.

4. THE EFFECT OF INTERLAMINAR DIFFUSION.

The modification of the distribution function consequent on eddy diffusion of the individual molecules in a radial direction may be computed by a method analogous to that used before (Bosworth 1948). We consider a hollow cylinder of internal radius r and external radius $r+dr$ entering the reaction zone. During passage through half the reaction zone a certain subfraction $f_x dx$ will diffuse from the hollow cylinder into another of radii x and $x+dx$. The probability that in time $\frac{1}{2}\tau$ a given molecule will diffuse a distance $\zeta (=x-r)$ is

$$A x dx e^{-\zeta^2/2 D_t \tau},$$

where the constant A is to be evaluated subject to the condition that the total quantity of material is not changed by diffusion, or $f_x dx$ integrated over all available space is unity.

As in the previous paper we will assume that the diverging cylindrical waves of diffusing particles are reflected at the walls and the converging cylindrical waves pass through the singularity at the centre and diverge again. For any particular value of x , ζ is a multivalued function $\zeta = x-r$, $x+r$, $2R-x-r$, $2R+x-r$, etc. Our sub-fraction f_x , using the notation of equation (19), now becomes

$$f_x dx = A x [e^{-\frac{b^2 R}{R-r}(x-r)^2} + e^{-\frac{b^2 R}{R-r}(x+r)^2} + \dots] dx.$$

The arguments of the exponentials in this expression can be regarded as small for the first two terms, so that the expansion

$$e^{-a^2} = 1 - a^2 + 0(a^4)$$

is permissible. The arguments of succeeding terms are larger and such terms may be neglected in comparison with the first two. Thus we have approximately

$$f_x = 2Ax \left[1 - \frac{b^2 R}{R-r} (x^2 + r^2) \right], \quad (20)$$

where A is given by the condition

$$\int_0^R f_x dx = 1$$

or

$$\begin{aligned} \frac{1}{A} &= \int_0^R 2x \left[1 - \frac{b^2 R}{R-r} (x^2 + r^2) \right] dx \\ &= R^2 - \frac{b^2 R}{R-r} \left(\frac{1}{2} R^4 + R^2 r^2 \right), \quad (21) \end{aligned}$$

which on substitution in equation (20) gives

$$f_x = \frac{2x}{R^2} \frac{1 - \frac{b^2 R}{R-r} (x^2 + r^2)}{1 - \frac{b^2 R}{R-r} \left(\frac{1}{2} R^2 + r^2 \right)},$$

or approximately

$$f_x = \frac{2x}{R^2} \left[1 + \frac{b^2 R}{R-r} \left(\frac{1}{2} R^2 - x^2 \right) \right]. \quad (22)$$

The average time of transit for a molecule which diffuses into the zone x to $x+dx$ in the time of half passage through the reactor will be taken as that appropriate to the statistical streamline point x . Therefore the subfraction $f_\tau d\tau$ which enters the hollow cylinder between r and $r+dr$ and leaves with a reaction time lying between τ and $\tau+d\tau$ is given by

$$f_\tau = 2n \frac{\tau_0^n}{\tau^{n+1}} \left(1 - \frac{\tau_0^n}{\tau^n} \right) \left[1 + \frac{b^2 R^3}{R-r} \left\{ \frac{1}{2} - \left(1 - \frac{\tau_0^n}{\tau^n} \right)^2 \right\} \right]. \quad . . . (23)$$

We shall use the symbol τ^\dagger to denote the quantity

$$\frac{1}{2} - \left(1 - \frac{\tau_0^n}{\tau^n} \right)^2,$$

τ^\dagger is thus a dimensionless quantity which is positive if

$$\sqrt{\frac{1}{2}} > 1 - \frac{\tau_0^n}{\tau_1^n},$$

or if

$$\tau < (3.41)^{1/n} \tau_0,$$

and attains its highest positive value of $\frac{1}{2}$ when $\tau = \tau_0$. (It is only when we come to deal with the effect of longitudinal diffusion that we have to consider the possibility of any value of τ less than τ_0 .) When

$$\tau > (3.41)^{1/n} \tau_0,$$

τ^\dagger assumes negative values and takes the (arithmetically) highest negative value of $-\frac{1}{2}$ when $\tau \rightarrow \infty$.

The fraction $F_\tau d\tau$ of the total molecular stream passing through the reaction zone with reaction time lying between τ and $\tau + d\tau$ is thus given by

$$F_\tau = \int_0^R F_\tau f_\tau dr,$$

or

$$\begin{aligned} F_\tau &= \frac{2(n+1)(2n+1)}{n R^{2+1/n}} \frac{\tau_0^n}{\tau^{n+1}} \left(1 - \frac{\tau_0^n}{\tau^n}\right) \int_0^R r(R-r)^{1/n} \left[1 + \tau^\dagger \frac{b^2 R^3}{R-\tau}\right] dr \\ &= 2n \frac{\tau_0^n}{\tau^{n+1}} \left(1 - \frac{\tau_0^n}{\tau^n}\right) [1 + (2n+1)b^2 R^2 \tau^\dagger]. \quad \dots \quad (24) \end{aligned}$$

Since values of τ less than τ_0 cannot result from purely radial diffusion it follows that F_τ must be zero when $\tau < \tau_0$ and thus equation (24) may be more completely written as

$$F_{\tau^*} = H(\tau^* - 1) 2n \tau^{*-n-1} (1 - \tau^{*-n}) [1 + (2n+1)b^2 R^2 \tau^\dagger], \quad \dots \quad (25)$$

where τ^* , as above, has been written for the dimensionless relative reaction time τ/τ_0 . Equation (25) compared with equation (11) shows the modification in the distribution function produced by radial turbulent diffusion. Apart from a slightly modified term in τ^* the distribution function involves only the dimensionless quantity bR which, it will be recalled, is equal to

$$(n-1) \sqrt{\frac{R}{0.32nL}},$$

and thus depends on the ratio of the radius of the reactor to its length and involves no other property.

5. THE EFFECT OF LONGITUDINAL DIFFUSION.

As in the former discussion on laminar motion (Bosworth 1948) the times of transit through the reaction zone for a particular molecule is modified if the molecule diffuses ahead (or behind) the initial statistical streamline point.

Consider any molecule for which the appropriate statistical streamline point has a transit time τ . The probability $p_z dz$ that a molecule originally occupying a point A on a statistical streamline will, in time τ , move ahead or behind A a distance lying between z and $z + dz$ is given by

$$p_z dz = \sqrt{\frac{b^2 R}{2\pi y}} e^{-b^2 z^2 R/2y} dz. \quad \dots \quad (26)$$

The quantity z is to be measured in the direction of flow. Accordingly the molecules which fall behind a distance z will have a transit time τ' given by

$$\tau' = \tau \frac{L}{L+z}.$$

The probability $p(\tau, \tau')d\tau'$ that a molecule travelling along a statistical streamline of transit time τ will itself have a transit time lying between τ' and $\tau'+d\tau'$ is given by

$$p(\tau, \tau') = \sqrt{\frac{b^2 \tau^n}{2\pi \tau_0^n}} \frac{L\tau}{\tau^2} e^{-b^2 L^2 \left(\frac{\tau}{\tau_0} - 1\right)^2 \tau^n / 2\tau_0^n}.$$

But the probability that the streamline point has a transit time lying between τ and $\tau+d\tau$ is given by equation (24) above. Accordingly the fraction of the total molecular stream having a transit time lying between τ' and $\tau'+d\tau'$ is $\Phi_{\tau'} d\tau'$ where

$$\Phi_{\tau'} = \frac{n}{\sqrt{\frac{1}{2}\pi}} [1 + (2n+1)b^2 R^2 \tau^n] \frac{bL}{\tau'^2} \left[\left(\frac{\tau_0}{\tau} \right)^{n/2} \left(1 - \frac{\tau_0^n}{\tau^n} \right) e^{-\frac{b^2 L^2}{2} \left(\frac{\tau}{\tau_0} - 1 \right)^2 \frac{\tau^n}{\tau_0^n}} d\tau. \right. \quad (27)$$

The integrand in equation (27) is that of a function which comes to a sharp maximum at some value of $\tau > \tau_0$. Accordingly the lower limit of integration could conveniently be made 0 instead of τ_0 without introducing a sensible error. Further $b^2 L^2$ is not a small quantity (it has a magnitude of order 300 or more) it follows that the exponential term in the integrand of equation (27) has a far sharper maximum than any power term. We may, therefore, without introducing any first-order error, remove any power term in τ from the integrand and replace it by a corresponding term outside the integral sign evaluated at that value of τ which makes the exponential term a maximum. The mode of the exponential term is clearly attained at

$$\tau = \tau',$$

and accordingly equation (27) may be written in the approximate form

$$\Phi_{\tau'} = \frac{n}{\sqrt{\frac{1}{2}\pi}} [1 + (2n+1)b^2 R^2 \tau^n] \frac{bL}{\tau'^2} \frac{\tau_0^n}{\tau'^n} \left(1 - \frac{\tau_0^n}{\tau'^n} \right) \int_0^\infty \frac{\tau^{n/2}}{\tau_0^{n/2}} e^{-\frac{b^2 L^2}{2} \left(\frac{\tau}{\tau'} - 1 \right)^2 \frac{\tau^n}{\tau_0^n}} d\tau. \quad (28)$$

The integrand in equation (28) may be written as the product of three exponentials; one in τ^{n+2} , one in τ^{n+1} and one in τ^n . The first varies rapidly in comparison with the other two which accordingly can be taken outside the integrand and evaluated at the mode. The integral in equation (28) therefore approximates to

$$\int_0^\infty \frac{\tau^{n/2}}{\tau_0^{n/2}} e^{-\frac{b^2 L^2}{2} \frac{\tau^{n+2}}{\tau'^{n+2}}} d\tau,$$

which is

$$\frac{2}{n+2} \frac{1}{\tau_0^{n/2}} \int_0^x e^{-\frac{b^2 L^2}{2} \frac{1}{\tau'^2 \tau_0^n} (\tau'^{n/2+1})^2} d(\tau'^{n/2+1}),$$

or

$$\frac{\sqrt{2\pi}}{n+2} \frac{\tau'}{bL}.$$

So that

$$\Phi_{\tau'} = \frac{2n}{n+2} [1 + (2n+1)b^2 R^2 \tau'^{\frac{n}{2}}] \frac{\tau_0^n}{\tau'^{n+1}} \left(1 - \frac{\tau_0^n}{\tau'^n}\right) e^{-\frac{b^2 L^2}{2} \left(\frac{\tau_0}{\tau'} - 1\right)^2}. \quad (29)$$

This equation, when compared with equation (10), shows the modification in the distribution function produced both by longitudinal and transverse eddy diffusion.

6. DISCUSSION.

Comparison of equations (29) and (10) show that, under the conditions for which the approximations used in the derivations above are valid, there is no instance for which we can ignore the effects of eddy diffusion which, it will be noted, depends on the two quantities $b^2 R^2$ and $b^2 L^2$, or on

$$(n-1)^2 \frac{R}{0.32nL} \quad \text{and} \quad (n-1)^2 \frac{L}{0.32nR}.$$

The effect of eddy diffusion on the shape of the distribution curve for reaction terms thus depends only on the ratio of length to diameter and on the value of n (which varies slightly with the Reynolds number).

Transverse diffusion produces the most profound modification of the distribution function when $\tau^{\frac{n}{2}}$ attains its maximum value which is $\frac{1}{2}$. The modifying factor involving $\tau^{\frac{n}{2}}$ then becomes

$$1 + \frac{1}{2}(2n+1)b^2 R^2.$$

The modification produced is less than 1 per cent only if

$$7.5 b^2 R^2 < 0.01$$

or

$$L/R > 12,000.$$

Under these conditions the longitudinal modification

$$e^{-\frac{1}{2}b^2 L^2 \left(\frac{\tau_0}{\tau'} - 1\right)^2},$$

becomes the dominating factor in the distribution curve which therefore practically takes the form of a "law of errors curve" with a standard deviation proportional to $1/bL$.

We may therefore summarize the findings of this paper by the statement that the distribution function for reaction times for a reagent passing turbulently through a cylindrical reactor consists of a sharply peaked curve (equation (29)) which exhibits a frequency maximum at a

reaction time only a little in excess of that calculated for the central core of fluid. The shape of the distribution curve is independent of all physical properties of the reagent or the reactor with the exception of the ratio of the length to the diameter, and to a minor degree, on the Reynolds' number and the roughness of the surface (which factors control the value of n in equation (29)). Under no readily attainable physical conditions can the condition of a distribution function practically independent of the ratio of length to diameter over a limited range be realized, as was the case when considering laminar flow.

REFERENCES.

- BOSWORTH, R. C. L., 1948, *Phil. Mag.* [7], **39**, 147.
 EWALD, P. P., PÖSCHL, T., and PRANDTL, L., 1930, *The Physics of Solids and Liquids*, p. 281 (London: Blackie and Sons).
 VON KARMAN, T., 1934, *J. of Aeronautical Sci.*, **1**, 1.
 PRANDTL, L., 1933, *V. D. I.* **7**, 5.
 SUTTON, O. G., 1934, *Proc. Roy. Soc. A*, [7], **146**, 701-22.

XXX. Change of Electrical Resistance of Alloys during Ageing.

By Z. MATYÁŠ,

The 1st Physical Institute of University of Technical Sciences in Prague*.

[Received August 10, 1948.]

ABSTRACT.

In the first part of this paper the author derives the mechanism by which the conduction electrons in pure aluminium are scattered when an external field is applied. Applying these results to dilute solid solutions of aluminium, it is possible to predict the change of the resistivity during ageing. Two alloys are considered here: duraluminium and aluminium-silver alloy in two different stages of ageing. In both cases the qualitative predictions of the theory are in agreement with experiment and even the quantitative results seem to be of the right order of magnitude.

1. INTRODUCTION.

DURING the age-hardening of supersaturated solid solutions the solute atoms separate as a separate phase. The application of X-ray methods to the examination of single crystals gives a fairly detailed picture of how the solute atoms move during the course of the reaction. Our present knowledge of ageing includes the following steps in the precipitation process (at least in the systems Al-Cu, Al-Ag and Al-Mg).

* Communicated by Prof. N. F. Mott.

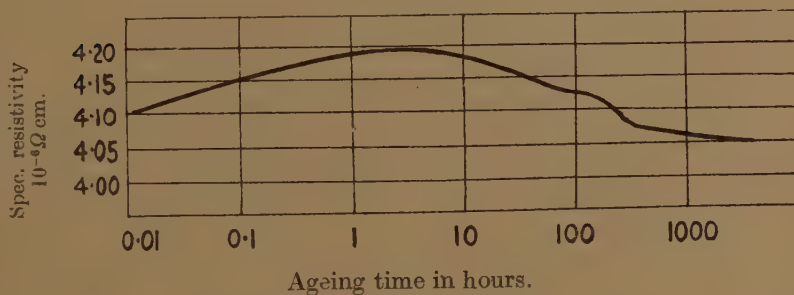
(a) Statistical fluctuations of concentration in the matrix provide clusters of atoms of the correct composition to form nuclei.

(b) In the early stages of precipitation these nuclei take the form of disks, grow rapidly and tend to take up a thin sheet-like form. In these plates, which are known as Guinier-Preston aggregates, the concentration of the solute atoms is higher than in the surrounding matrix and they are segregated on definite planes of the crystal (*e. g.* on (100) planes in duraluminium).

(c) The particles grow in thickness and in lateral dimensions, and finally in some alloys attains a size that yields three-dimensional X-ray diffraction.

(d) Upon reaching a certain size the transition structure is transformed to the stable structure of the equilibrium precipitate.

Fig. 1.



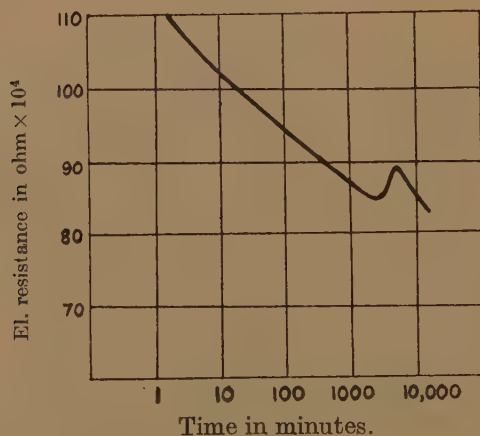
This transition to the state of equilibrium manifests itself in many ways. A change in hardness, in electrical resistivity, in the Curie temperature and in the intensity of magnetization during ageing is observed.

The present paper is entirely concerned with the change of resistivity of binary Al-Cu or Al-Ag alloys that takes place during the precipitation of copper or silver respectively. In fig. 1 we show the variation with time of the resistivity of a quenched aluminium-copper alloy at room temperature (Fink 1942). The copper content of the alloy is 4 per cent. It may be seen that the resistance passes through a maximum during the precipitation process and then drops to a value lower than that for the case in which the copper is dissolved in the lattice. Fig. 2 shows the change in electrical resistance with time on ageing an aluminium-silver alloy at the temperature 158°C. (the silver content is 20 per cent) (Geissler, Barret and Mehl 1943). The electrical resistance in this case decreases immediately from the first measurement and continues to decrease in a uniform manner during ageing for a considerable time. This initial decrease is then followed by a rise of resistance to an intermediate maximum.

The origin of the change in resistivity during precipitation is probably twofold. In the first place, the lattice strain which accompanies the formation of nuclei can influence the resistivity. The flow of electrons through a lattice is sensitive to any disturbance in the periodicity of the lattice. The degree of internal strain, especially when the nuclei are small, can increase the perturbation from the periodic potential in which the conduction electrons in the perfect lattice move; this gives rise to a scattering probability of electrons and the resistivity of the alloy must increase (fig. 1). But as time goes on the particles tend to take up a thin sheet-like form, the strain diminishes and we can expect a uniform decrease of resistance.

As Mott (1937) first pointed out, however, this cannot be the only factor influencing the rise, for the peaks for hardness and resistivity do not occur at the same time, as we should expect (if they had the same

Fig. 2.



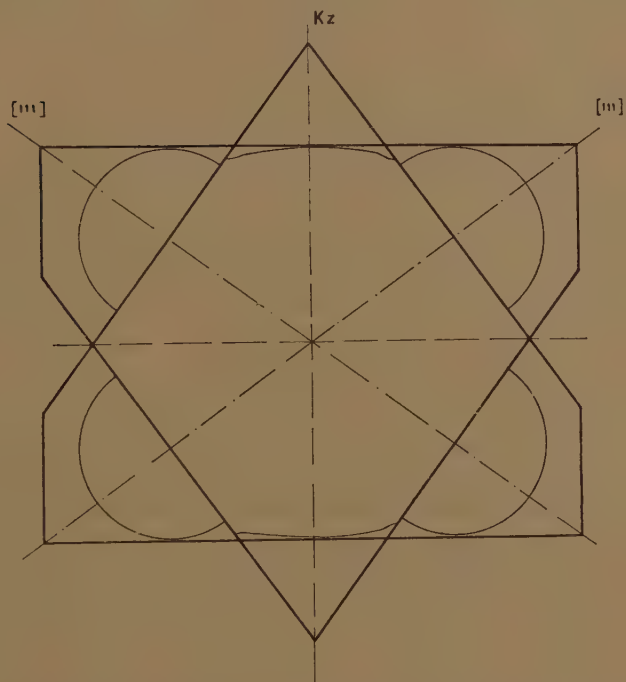
cause). Mott proposed that the peak value of the resistivity occurs when the precipitated particles attain the size for which the scattering of electrons is a maximum. On this basis the "abnormal" resistivity, observed in the early state of ageing (fig. 1), should decrease as the nuclear size becomes larger.

In our opinion the change of scattering power of the nuclei of different size must be the only factor determining the resistivity change in systems where the size difference of the atoms is small (*i. e.* in our systems Al-Cu or Al-Ag). As we have already obtained in a previous paper (Matyáš 1948) the character of the Brillouin zones for aluminium, it will be possible to construct a fairly detailed picture of the scattering mechanism of the conduction electrons due to foreign atoms present in a dilute solid solution in this metal. On this basis we will be able to carry out approximate calculations of the scattering power of Guinier-Preston aggregates of different size and thus derive a formula for the relative change in the specific electrical resistivity during the ageing of the alloy.

2. ENERGY BAND STRUCTURE IN ALUMINIUM

The three valence electrons of aluminium are distributed in the first and second Brillouin zones. A face-centred cubic lattice of this metal has a body-centred lattice as its reciprocal, and it is easy to show that the first Brillouin zone is a truncated octahedron, marked out by the planes (111) and (200). The second zone is bounded by the same set of planes. Fig. 3 represents the section through these zones in the plane containing the axis k_z and the direction (111) respectively, together with the position of the Fermi energy surfaces, as derived in the paper mentioned above. The first zone is approximately full. It contains,

Fig. 3.



therefore, nearly two electrons and the Fermi surface lies near to the surface faces in the (100)-directions. There is no overlap in the (100) direction into the second zone. The third electron is situated in the second zone where the Fermi surface has the shape of semi-ellipsoids, surrounding the eight points $(\pm\pi/a, \pm\pi/a, \pm\pi/a)$.

To make the calculation manageable and yet to obtain the essential features of the scattering mechanism of the conduction electrons in aluminium, we shall approximate these semi-ellipsoids by spheres. We shall therefore assume that the energy of the conduction electrons at the Fermi surface in the (111)-directions is given by the formula :

$$E = E_0 + \hbar^2 k^2 / 2m, \quad \dots \dots \dots (1)$$

where E_0 means the lowest energy of the second band, which is according to our calculation (Matyáš 1948) equal to 9.5 eV. It is well known (Mott and Jones 1936) that the number of electrons per atom in a volume V of k space (atomic volume $\tau = a^3/4$) is given by

$$n = 2(\tau/8\pi^3)V = (a^3/16\pi^3)V. \quad (2)$$

The third electron must be distributed in the second zone, where, according to our assumption, the energy contour of the Fermi limit has the shape of semi-spheres, surrounding the eight points $(\pm\pi/a, \pm\pi/a, \pm\pi/a)$. The corresponding volume V is thus equal to

$$V = 4 \cdot \frac{4}{3}\pi k_m^3. \quad (3)$$

From equations (2) and (3) we can now determine the maximum wave vector k_m

$$k_m = |\mathbf{k}_m| = \pi/a.$$

The Fermi energy from equation (1) is equal to

$$E_F = 9.5 + \hbar^2(\pi/a)^2/2m = 9.5 + 2.5 = 12 \text{ eV.}, \quad (4)$$

which is in agreement with Skinner's measurements of the soft X-ray emission band (12.0 eV.). At the same time we see from (1) that the corresponding wave functions of the conduction electrons with the Fermi energy are plane waves :

$$\psi(\mathbf{k}_m) = \exp \{i(\mathbf{k}_m \mathbf{r})\}. \quad (5)$$

Further, we know that the situation near to the square faces (200) in the first zone is as follows :—The contours of the Fermi energy are practically parallel to the bounding planes (200) and only near to the corners of the first zone are they slightly curved from this plane (fig. 3). The high density of these states is closely connected with this fact.

For further considerations we need to know the wave functions of the states which correspond to the Fermi surface parallel to the square faces. The energy of these states can be expressed by the formula (Matyáš 1948)

$$E' = E'_0 - 10\hbar^2 j_n^2/2m, \quad (6)$$

where E'_0 denotes the highest energy in the first zone, and j_m is the vector in the direction (200), with magnitude equal to $2\pi/a - k'_{mx}$, k'_{mx} being the x -component of the wave vector of these states (fig. 3). As we know that k'_{mx} is only slightly different from $k_0 = 2\pi/a$, we must conclude that $j_m \ll \pi/a$. According to the expression (6) the wave functions of the states near to the mid-point of the square faces (200) have the form

$$\psi(\mathbf{k}'_{mx}) = \exp \{i(\mathbf{j}_n \mathbf{r})\}. \quad (7)$$

3. CONDUCTION OF ELECTRICITY IN ALUMINIUM

We have seen in the preceding paragraph that the first zone is nearly completely filled. The number of positive holes in this zone is very small and only the electrons in the second zone contribute to the conductivity.

These conclusions are probably over simplified, but they will be sufficient for our purposes.

If we thus apply an external electrical field only the electrons of the second zone can be accelerated. Now the time of relaxation τ_p for these electrons is given by

$$1/\tau_p = P,$$

where P is the probability of scattering. An electron in the second zone can be scattered by perturbations from a perfect periodic potential in two ways: either to another state in the second zone or to an s state in the first zone.

By the ordinary methods of quantum mechanics (Mott and Jones 1936), one finds that if the electron is initially in the state \mathbf{k} , the probability per unit time of a transition to the state \mathbf{k}' is

$$P(\mathbf{k}, \mathbf{k}') dS = \frac{dS}{4\pi^2 \hbar} \left| \int \psi(\mathbf{k}') \Delta V \psi(\mathbf{k}) d\tau \right|^2 \left/ \frac{dE}{dk_n} \right., \quad \dots \quad (8)$$

where dS means the element of Fermi surface in k -space, dE/dk_n the differentiation of energy in the direction normal to the Fermi surface, ΔV is the perturbing potential and $\psi(k)$ or $\psi(k')$ the wave functions of the initial or final state respectively.

From this formula and equation (5) or (7) we can derive the expression for the probability of a transition between two states in the second zone:

$$\begin{aligned} P(k_m, k'_m) &= \frac{dS}{4\pi^2 \hbar} \left| \int \exp \{i(k_m - k'_m) \cdot r\} \Delta V d\tau \right|^2 \left/ \frac{dE}{dk_m} \right. \\ &= \frac{dS}{4\pi^2 \hbar} \left| \iiint \exp (2ik_m r \sin \frac{1}{2}\theta \cos \delta) \Delta V r^2 \sin^2 \delta r d\delta d\phi \right|^2 \left/ \frac{dE}{dk_m} \right. \\ &\dots \dots \dots (9a) \end{aligned}$$

and the probability of a transition between the states in the second zone and the states lying in the first zone characterized by the wave function (7):

$$\begin{aligned} P(k_m, j) dS &= \frac{dS_j}{4\pi^2 \hbar} \left| \iiint \exp (ik_m r^2 \cos \delta) \Delta V r^2 \sin \delta dr d\delta d\phi \right|^2 \left/ \frac{dE'}{dj_n} \right.^* \\ &\dots \dots \dots (9b) \end{aligned}$$

We see from these formulæ that the transition probabilities are proportional to dS/dk_m or dS_j/dj_n , and hence to the density of states at the Fermi surface. If we introduce polar coordinates with the pole at the origin of k -space, then from equations (1) and (6) we can write

$$dS \left/ \frac{dE}{dk_m} \right. = \frac{k_m^2 \sin \delta d\delta d\phi}{k_m \hbar^2 / m} = \frac{\pi \sin \delta d\delta d\phi}{a \hbar^2 / m}, \quad \dots \quad (10a)$$

* We take into account that $j_n < k_m$.

and

$$dS_j / \frac{dE'}{dj_n} = \left(\frac{2\pi}{a} \right)^2 \frac{\sin \delta \, d\psi \, d\delta}{10j_n \hbar^2 / m} \dots \dots \dots (10b)$$

Therefore the probability $P(\mathbf{k}_m, \mathbf{k}'_m)$ is much smaller than $P(\mathbf{k}_m, \mathbf{j}_n)$.

Similarly we can show that the probability of a transition from a state in the second zone to the states lying on the Fermi contour near the corner of the first zone is negligibly small compared with $P(k_m, j_n)$ (here the energy can be expressed by the formula $E = E'_0 - \alpha \hbar^2 j_n^2 / 2m$, where $j'_n \ll \pi/a$).

We obtain the following picture of electrical conductivity in aluminium. The current is carried by one electron lying in the second zone. The resistance is mainly due to scattering processes in which the electron makes a transition from the states of the second zone to s -states in the first zone lying on the Fermi distribution near the bounding square faces of the first Brillouin zone.

We must now derive the formula for the electrical resistivity and hence give a more exact definition of τ_p in terms of the transition probability (9b).

Consider the metal in the absence of a field. The probability that any electronic state is occupied is given by the Fermi function

$$f(\mathbf{k}_m) = f_0(\mathbf{k}_m) = 1 / (e^{(E - \zeta)/kT} + 1), \quad E = E(\mathbf{k}). \quad \dots \dots (11)$$

If an external field \mathbf{F} is applied, say in the x -direction, the wave vector of each electron in the second zone increases according to the equation $k_x = -eF/\hbar$; hence the rate of change of $f(\mathbf{k}_m)$ is initially

$$\begin{aligned} \left(\frac{df}{dt} \right)_F &= - \frac{df_0}{dE} \frac{dE}{dk} \frac{k_{mx}}{k_m} \frac{eF}{\hbar} \\ &= - \frac{df_0}{dE} \frac{\hbar}{m} k_{mx} eF. \quad \dots \dots \dots (12) \end{aligned}$$

This expression must be equated to the rate of change of f due to collisions.

When a steady current is flowing, the probability that a state is occupied will no longer be given by $f_0(k_m)$ but by the "perturbed" Fermi functions $f_p(k_m)$, $f_s(j_n)$ for electrons in the second or first zone respectively. We shall write

$$\left. \begin{aligned} f_p(\mathbf{k}_m) &= f_0(\mathbf{k}_m) + g_p(\mathbf{k}_m) \\ f_s(\mathbf{j}_n) &= f_0(\mathbf{j}_n) + g_s(\mathbf{j}_n) \end{aligned} \right\} \dots \dots \dots (13)$$

By the definition of the time of relaxation τ_p , we have (according to equation (12))

$$g_p = \frac{df}{dt} \tau_p = - \frac{df_0}{dE} \frac{\hbar}{m} k_{mx} eF \tau_p, \quad \dots \dots \dots (14)$$

which gives the change of the distribution function due to the field \mathbf{F} , in terms of the time of relaxation τ_p .

Similarly we obtain for g_s the expression

$$g_s = \frac{df}{dt} \tau_s = - \frac{df_0}{dE} \frac{\hbar}{m} 10j_{nx} e F \tau_s. \quad (15)$$

We now calculate the rate of change of f due to collisions, which has to be equated to (12).

Consider a volume element of k space in the second zone. Let the volume element be situated at the point \mathbf{k} and let its volume be $d\mathbf{k} = dk_x dk_y dk_z$. Then for unit volume of the metal the number of electrons in states within the element $d\mathbf{k}_m$ is

$$2f_p(\mathbf{k}_m) d\mathbf{k}_m / 8\pi^3.$$

Owing to the perturbation potential, electrons make transitions to states of practically equal energy in the first zone; the number of such transitions in which an electron leaves the volume element $d\mathbf{k}_m$ at the Fermi surface in the second zone is

$$2f_p(\mathbf{k}_m) \frac{d\mathbf{k}_m}{8\pi^3} \int P(\mathbf{k}_m, \mathbf{j}_n) \{1 - f_s(\mathbf{j}_n)\} dS_j. \quad (16)$$

The integration is over all surface elements dS_j in j -space which have the same energy as the initial state and lie on the Fermi surface parallel to the bounding square faces in the first zone. If we introduce the notation

$$\left. \begin{aligned} P(\mathbf{k}_m, \mathbf{j}_n) dS_j &= \frac{1}{10} \frac{dS_j}{j_n} A(\mathbf{k}_m, \mathbf{j}_n) \\ A(\mathbf{k}_m, \mathbf{j}_n) &= \frac{m}{4\pi^2 \hbar^3} \left| \int \exp(i\mathbf{k}_m \mathbf{r} \cos \delta) \Delta V d\tau \right|^2 \end{aligned} \right\} \quad (17)$$

we can write the expression (16) in the form

$$2f_p(\mathbf{k}_m) \frac{d\mathbf{k}_m}{8\pi^3} \frac{1}{10j_n} \int A(\mathbf{k}_m, \mathbf{j}_n) \{1 - f_s(\mathbf{j}_n)\} dS_j. \quad (18)$$

Similarly, we obtain for the number of electrons entering the volume element $d\mathbf{k}_m$ in the second zone

$$2\{1 - f_p(\mathbf{k}_m)\} \frac{d\mathbf{k}_m}{8\pi^3} \frac{1}{10j_n} \int A(\mathbf{k}_m, \mathbf{j}_n) f_s(j) dS_j. \quad (19)$$

Subtracting (19) from (18) we obtain for the decrease, due to collisions, in the number of electrons in the volume element per unit time

$$2 \frac{d\mathbf{k}_m}{8\pi^3} \frac{1}{10j_n} \int A(\mathbf{k}_m, \mathbf{j}_n) \{f_p(\mathbf{k}_m) - f_s(\mathbf{j}_n)\} dS_j. \quad (20)$$

This quantity (20) vanishes if we substitute the unperturbed Fermi function f_0 for f_p and f_s . Therefore, in formula (20) we may substitute g_s for f_s and g_p for f_p . Using (14) and (15) we obtain

$$-2 \frac{d\mathbf{k}_m}{8\pi^3} \frac{1}{10j_n} \frac{df_0}{dE} \frac{\hbar}{m} e F \int A(\mathbf{k}_m, \mathbf{j}_n) \{\tau_p k_{mx} - 10j_{nx} \tau_s\} dS_j. \quad (21)$$

In the last expression the integral

$$\int A(\mathbf{k}_m, \mathbf{j}_n) 10 j_{nx} \tau_s dS_j = A(\mathbf{k}_m, \mathbf{j}_n) 10 \tau_s \int j_{nx} dS_j$$

vanishes on account of the symmetry and we can thus write instead of (21)

$$-2 \frac{dk_m}{8\pi^3} \frac{1}{10j_n} \frac{df_0}{dE} \frac{\hbar}{m} eF \tau_p \int A(\mathbf{k}_m, \mathbf{j}_n) k_{mx} dS_j. \quad (22)$$

This must be equated to the increase in the number of electrons due to the applied field, which is, from (12),

$$\frac{2dk_m}{8\pi^3} \left(\frac{df}{dt} \right)_F = -2 \frac{dk_m}{8\pi^3} \frac{df_0}{dE} \frac{\hbar}{m} eF k_{mx}. \quad (23)$$

Equating (23) and (22) and dividing by common factors we get

$$\frac{1}{\tau_p} = \frac{1}{10j_n} \int A(\mathbf{k}_m, \mathbf{j}_n) dS_j. \quad (24)$$

The current due to the electrons in the second zone is

$$i = e \int g_p(\mathbf{k}_m) v_x dk_m / 8\pi^3,$$

where $v_x = \hbar k_x / m$ is the velocity of a p electron. If n_p is the number of electrons in the second zone, this easily reduces to

$$i = - \frac{n_p e^2 F}{m} \int \tau_p \frac{df_0}{dE} dE. \quad (25)$$

Since df_0/dE is a function which vanishes everywhere except in the neighbourhood of $E = E_m$, i. e. at the surface of the Fermi distribution, and since further $-\int (df_0/dE) dE = 1$, this reduces to

$$i = n_p e^2 F \tau_p / m. \quad (26)$$

Therefore the specific resistivity is given by the formula

$$\rho = m / n_p e^2 \tau_p. \quad (27)$$

4. CHANGE OF RESISTIVITY DURING THE AGEING.

After these considerations we can calculate the change of the specific resistivity of dilute solid solutions of copper or silver in aluminium during precipitation. First of all, we shall make the following assumption:—As we are interested in aluminium alloys, in which the foreign atoms are present in very small concentration, the mechanism of electrical conductivity is exactly the same as we have described in the preceding paragraph.

We know further that in a pure metal the only deviations from the periodic potential are due to the thermal vibrations, which give rise to a scattering probability P_t proportional to the temperature. In solid solutions there will be a further departure of the field from periodicity in the places which are occupied by foreign atoms, and hence there will

be a further scattering probability P_0 independent of the temperature. It may be shown that these probabilities are independent of one another; we can thus write formula (27) in the form

$$\rho = m(P_t + P_0)/n_p e^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

and hence the term which governs the change of resistivity of an alloy during ageing at constant temperature is given (equations (17) and (24)) by $mP_0/e^2 n_p$, where

$$P_0 = \frac{1}{10j_n} \int dS_j \frac{m}{4\pi^2 \hbar^3} \left| \int \exp(i\mathbf{k}_m \mathbf{r}) \Delta V d\tau \right|^2, \quad . \quad . \quad . \quad (29)$$

ΔV being the perturbing potential due to the presence of foreign atoms.

We shall here confine ourselves to two cases. First we shall consider the *earliest stage* of ageing, when some foreign atoms come together and form the nuclei, which grow in size. We shall assume that these nuclei have spherical form and that the perturbing potential ΔV in the expression (29) inside the single foreign atoms and in the clusters is approximately constant.

From (28) and (29) we can write the formula for the specific resistivity during ageing in the form:

$$\rho = c_1 + c_2 \left| \int \exp(i\mathbf{k}_m \mathbf{r}) \Delta V d\tau \right|^2, \quad . \quad . \quad . \quad . \quad . \quad (30)$$

where c_1 and c_2 are constants during precipitation. We can split the perturbation potential into parts: $\Delta V = \Delta V_1 + \dots + \Delta V_n$; ΔV_n is equal to a constant value A inside the sphere which replaces a foreign atom or cluster of atoms and is equal to zero outside this sphere. Accordingly, we can split our integral in (30) into parts; if we introduce polar coordinates r, δ, ψ , with the polar axis along k_m and with the pole at the centre of each foreign particle, we obtain

$$\rho = c_1 + c_2 A^2 \left| \sum_p \exp(i\mathbf{k}_m \mathbf{r}_p) \int \exp(i\mathbf{k}_m \mathbf{r}) d\tau_p \right|^2, \quad . \quad . \quad (31)$$

where r_p means the lattice vector giving the position of a foreign atom or the nuclei respectively and $\int d\tau_p$ denotes the integral over the volume of a particle.

If we carry out the integration in (31) for a particle of radius r , we obtain

$$\zeta_p = \int \exp(i\mathbf{k}_m \mathbf{r}) d\tau_p,$$

which reduces to

$$\zeta_p = \Omega \phi(k_m r), \quad . \quad . \quad . \quad . \quad . \quad (32)$$

where $\phi(x)$ denotes the function

$$\phi(x) = 3x^{-3}(\sin x - x \cos x), \quad . \quad . \quad . \quad . \quad . \quad (33)$$

and Ω the volume of a particle or cluster of foreign atoms.

Introducing these integrals (32) into (31) we can write

$$\begin{aligned} \rho &= c_1 + A^2 c_2 \left| \sum_p \zeta_p \exp(i\mathbf{k}_m \mathbf{r}) \right|^2 \\ &= c_1 + A^2 c_2 \sum_p \sum_{p'} \zeta_p \zeta_{p'} \exp\{i\mathbf{k}_m (\mathbf{r}_p - \mathbf{r}_{p'})\}. \quad . \quad . \quad . \quad (34) \end{aligned}$$

Assuming that the total number of foreign atoms in the alloy is equal to N and the number of nuclei is n , the number of single foreign atoms present in the alloy in a certain stage of ageing is given by $N - n\Omega/\Omega_0$, where $\Omega_0 = 4/3\pi r_0^3$ is the volume of a foreign atom. The double summation in (34) contains two types of integrals, ζ_0 and ζ_i ; the first corresponds to a single foreign atom and the second to a cluster of atoms. Therefore, we can write the expression for the resistivity (34) more explicitly:

$$\rho = c_1 + A^2 c_2 \{ (N - n\Omega/\Omega_0) \zeta_0^2 + n \zeta^2 + \sum_p \sum_{p'} \exp \{ ik_m(r_p - r_{p'}) \} \zeta_p \zeta_{p'} \}. \quad (35)$$

We denote by $r_n, r_{n'}$ the position vectors of the single foreign atoms and by $r_m, r_{m'}$ the position vectors of the clusters. The double summation in the last formula can be expressed as follows:

$$\begin{aligned} \sum_p \sum_{p'} \exp \{ ik_m(r_p - r_{p'}) \} \zeta_p \zeta_{p'} &= \zeta_0^2 \sum_n \sum_{n'} \exp \{ ik_m(r_n - r_{n'}) \} \\ &+ \zeta^2 \sum_m \sum_{m'} \exp \{ ik_m(r_m - r_{m'}) \} + \zeta_0 \zeta \left[\sum_m \sum_n \exp \{ ik_m(r_n - r_m) \} \right. \\ &\left. + \sum_n \sum_m \exp \{ ik_m(r_m - r_n) \} \right]. \end{aligned}$$

It is now possible to calculate the value of the terms on the right-hand side of the last expression. The first term is equal to the mean value of $\exp \{ ik_m(r_n - r_{n'}) \}$ times the number of terms in the sum. The evaluation of this mean value is rather easy, if we consider that the orientation of the crystallites in the alloy is at random; it has already been done by Debye (1927). We obtain

$$\begin{aligned} \sum_n \sum_{n'} \exp \{ ik_m(r_n - r_{n'}) \} &= - \left(N - n \frac{\Omega}{\Omega_0} \right) \left(n - 1 - \frac{n\Omega}{\Omega_0} \right) \frac{\Omega_0}{V} \phi(2k_m r_0) \\ \sum_m \sum_{m'} \exp \{ ik_m(r_m - r_{m'}) \} &= -n(n-1) (\Omega/V) \phi(2k_m r) \\ \sum_m \sum_n \exp \{ ik_m(r_n - r_m) \} &= -n(N - n\Omega/\Omega_0) (\Omega/V) \psi(2k_m r) \\ \sum_n \sum_m \exp \{ ik_m(r_m - r_n) \} &= -n(N - n\Omega/\Omega_0) (\Omega_0/V) \psi(2k_m r_0). \end{aligned}$$

Introducing these values into (35) we see at once that their order of magnitude is very small compared to $(N - n\Omega/\Omega_0) \zeta_0^2 + n \zeta^2$ and we can neglect them.

The final expression for the electric resistivity of an alloy is then of the form

$$\begin{aligned} \rho &= c_1 + A^2 c_2 \{ (N - n\Omega/\Omega_0) \zeta_0^2 + n \zeta^2 \} \\ &= c_1 + A^2 c_2 \{ (N - n\Omega/\Omega_0) \Omega_0^2 \psi^2(k_m r_0) + n^2 \Omega^2 \psi^2(k_m r) \}. \quad (36) \end{aligned}$$

From this we get the expression for the resistivity immediately after quenching, when we put $\Omega = \Omega_0$:

$$\rho_0 = c_1 + A^2 c_2 N \Omega_0^2 \phi(k_m r_0), \quad (37)$$

and

$$\overline{\zeta^2} = p^2 \zeta^2, \quad (41)$$

where ζ is expressed by formula (32).

Introducing this value into (36) we obtain for the specific resistance of the alloy

$$\rho = c_1 + A^2 c_1 \{ (N - pn\Omega/\Omega_0) \zeta_0^2 + np^2 \zeta^2 \}. \quad . . . (42)$$

Hence for the relative change of the resistivity we have

$$\frac{\rho - \rho_0}{\rho_0 - \rho'_0} = \frac{n}{N} px^3 \left\{ px^3 \frac{\psi^2(k_m r_0 x)}{\psi^2(k_m r_0)} - 1 \right\}. \quad . . . (43)$$

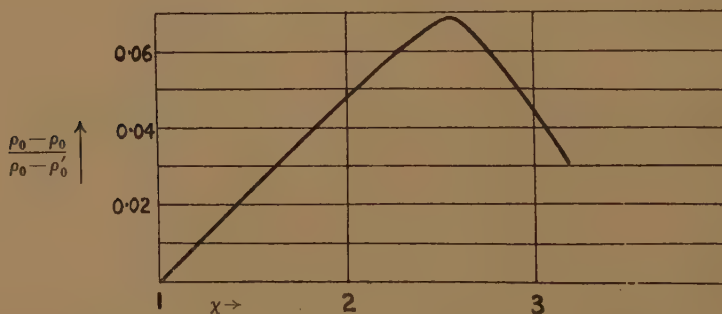
As in this case the radii of the clusters are large, the value of $\psi(k_m r_0 x)$ is practically zero and the last formula reduces to

$$\frac{\rho - \rho_0}{\rho_0 - \rho'_0} = -\frac{n}{N} px^3. \quad . . . (44)$$

5. COMPARISON WITH THE EXPERIMENTAL RESULTS.

To compare the theoretical formula (40) which describes the relative change of the electrical resistivity in the early stage of ageing with experiment, we shall use the measurements on the aluminium-copper alloy (fig. 1). We know already from the second paragraph that the

Fig. 4.



absolute magnitude of the wave vector k_m corresponding to the Fermi energy in the second zone is equal to π/a . The radius of a copper atom is $r_0 = 1.4 \times 10^{-8}$ cm. Now from the measurements of Guinier we know that the average distance between the nuclei immediately after quenching is about 70 Å. Therefore the ratio n/N in the Al-Cu alloy with 4 per cent of copper is equal to 2.7×10^{-3} . Putting these values into (40) and plotting the dependence of the relative change of resistivity upon $x = r/r_0$, we obtain the curve in fig. 4. We see at once that the predictions are in qualitative agreement with the experimental results (fig. 1). The resistance increases immediately after the quenching, passes through a maximum and then decreases steadily. According to our theoretical curve, the value of the maximum of $(\rho - \rho_0)/(\rho_0 - \rho'_0)$ is 0.068 and from the experimental results we obtain nearly the same value 0.067.

Further comparison between experiment and theory is not possible, as we know nothing about the growth of nuclei during precipitation. Also, we do not expect that the decrease of the resistivity is as steep as our curve predicts. It is well known that in the latter stage of ageing the particles are not spherical, but tend to take up a thin sheet-like form. The scattering of the conduction electrons in this case is not reproduced by our formula (40).

On the other hand, in the stage of the precipitation where the particles are rather thick we can again represent them by the spherical clusters and we can use formula (44). This case we can find in the alloy Al-Ag, with 20 per cent of silver (fig. 2). As the ageing temperature is rather high (158° C.) we must expect big particles; according to Guinier (1942) the radius of the clusters after one hour of ageing is about 50 Å. We can observe from fig. 2 a rather steep decrease of resistivity, which is predicted by our formula (44). The resistance ρ'_0 corresponding to pure aluminium is in our case equal to $55 \cdot 10^{-4}$ ohm, and hence the relative decrease of resistivity after one hour of ageing in this measurement is 0.25. This can be compared with the theoretical predictions according to formula (44). The distance between nuclei after one hour of ageing is about 140 Å (Guinier 1942) and therefore n/N in the expression (44) is equal to 10^{-4} . The density of the foreign atoms in the particles (p) is about 1/3 in this case. If we introduce these values into (44) we obtain for $x=25$ A/1.5 A = 16.7 $(\rho - \rho_0)/(\rho_0 - \rho'_0) = 0.16$.

The reader may find further discussion concerning the curve of fig. 2 in the paper of Geissler, Barrett and Mehl (1943).

The maximum in the resistivity in the early stage of ageing is caused by the peculiar mechanism of the scattering of the conduction electrons. It is possible to show that in those cases in which the Fermi limit lies only in the first zone, we must expect a decrease of resistance immediately after quenching. No maximum is observed in the resistivity curve, e. g., in the alloy Ag-Cu, where we are certain that the Fermi energy lies in the first zone.

I have great pleasure in acknowledging my indebtedness to Professor N. F. Mott for helpful interest in this work and for his advice in discussion on many occasions.

REFERENCES.

- DEBYE, 1927, *Phys. Zeit.*, **27**, 137-8.
 FINK, 1942, *J. Applied Phys.*, **13**, 75.
 GEISSLER, BARRETT, and MEHL, 1943, *A.I.M.E. Techn. Publication* No. 1557.
 GUINIER, 1942, *Journ. de Physique*, Series III, 133.
 MATYÁŠ, 1948, *Phil. Mag.* [7], **39**, 429.
 MOTT, discussion of Gayler, 1937, *Journal of the Inst. of Metals*, **60**, 267.
 MOTT and JONES, 1936, *Properties of Metals and Alloys* (Oxford University Press), p. 56, eq. 25; *op. cit.*, pp. 250-251.

XXXI. *The Flow of Viscous Fluids round Plane Obstacles.*

By R. ROSCOE, Ph.D.,

Physics Department, King's College, Newcastle-upon-Tyne*.

[Received March 19, 1948.]

ABSTRACT.

In certain cases of slow viscous flow round thin plates the solution of Stokes' equations can be reduced to a solution of Laplace's equation with simple boundary conditions, the flow being thus relatable to the potential distribution in an electrostatic position. It is shown that the force on a plate held perpendicular to a stream of velocity U is $8\pi\mu CU$, μ being the viscosity of the fluid and C the electrical capacity of a conductor of the same shape as the plate. The rate of flow of fluid through an elliptic aperture in a thin wall is shown to be $2S^2/3\pi\mu s$ times the pressure difference between the two sides, S being the area of the aperture and s its perimeter. A simple expression is also derived for the torque on an elliptic plate rotating about its minor axis. Approximate methods for dealing with the flow past plates of complicated shapes are described, and the case of a paddle-type viscometer is treated as an example.

§ 1. A SOLUTION OF STOKES' EQUATIONS.

FOR an incompressible fluid, Stokes' equations for slow viscous flow (Lamb 1924a) and the equation of continuity may be written as

$$\left. \begin{aligned} \mu \nabla^2 u &= \frac{\partial p}{\partial x}, & \mu \nabla^2 v &= \frac{\partial p}{\partial y}, & \mu \nabla^2 w &= \frac{\partial p}{\partial z} \end{aligned} \right\}, \quad \dots \quad (1)$$

$$\nabla^2 = \partial^2/\partial x^2 + \partial^2/\partial y^2 + \partial^2/\partial z^2$$

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0, \quad \dots \quad (2)$$

where μ is the coefficient of viscosity, u, v, w are the velocity components parallel to the x, y, z axes respectively, and p is the mean pressure in the fluid. These equations taken together with the boundary conditions suffice to determine the motion of the fluid. The normal tensions p_{xx} ,

* Communicated by Prof. W. E. Curtis, D.Sc., F.R.S.

p_{yy} , p_{zz} , at any point may be obtained from the relations

$$\left. \begin{aligned} p_{xx} &= -p + 2\mu \frac{\partial u}{\partial x}, \\ p_{yy} &= -p + 2\mu \frac{\partial v}{\partial y}, \\ p_{zz} &= -p + 2\mu \frac{\partial w}{\partial z}. \end{aligned} \right\} \dots \dots \dots (3)$$

Now if ϕ be any function obeying Laplace's equation $\nabla^2\phi=0$, the continuity equation (2) is satisfied by the flow system

$$\left. \begin{aligned} u &= \phi - x \frac{\partial \phi}{\partial x}, \\ v &= -x \frac{\partial \phi}{\partial y}, \\ w &= -x \frac{\partial \phi}{\partial z}. \end{aligned} \right\} \dots \dots \dots (4)$$

Also, on substituting the velocity components given by (4) in the left-hand members of (1), it is found that the flow system (4) is a solution of Stokes' equations with

$$p = -2\mu \frac{\partial \phi}{\partial x} \dots \dots \dots (4a)$$

This simple result forms the basis of the methods of calculation developed in the present work.

Let ϕ be taken as the electrical potential round an earthed, conducting, plane lamina so that ϕ is zero at the surface of the lamina, and let the origin of coordinates be taken in the plane of the lamina with the x axis normal to its plane. Now the electric field at a point near to a conducting lamina approaches a finite limit as the point moves towards a surface of the lamina, while if the point moves towards the edge the field increases inversely as the square root of the radial distance from the edge; so in either case the quantities $x\partial\phi/\partial x$, $x\partial\phi/\partial y$, $x\partial\phi/\partial z$ tend to vanish in the limit as the point approaches the lamina. Under these circumstances, the values of u , v , and w given by (4) vanish at the surfaces and edge of the lamina, so that the flow system represents a state of motion of a viscous fluid round a thin plate of the same form as the conducting lamina. As an example, fig. 1*a* shows the potential distribution round the mid-section of an earthed lamina whose length is much greater than its breadth, and fig. 1*b* shows the stream lines of the derived flow system.

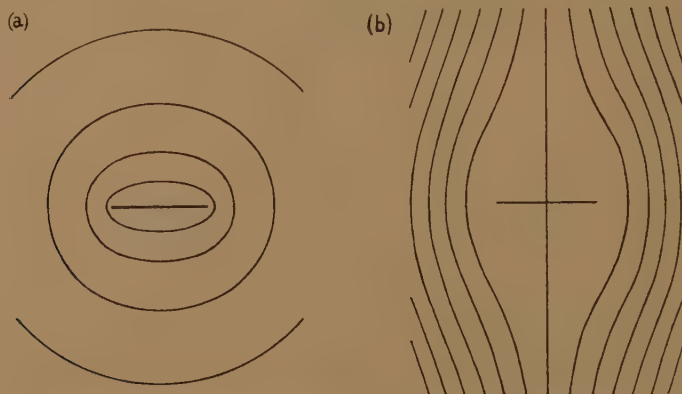
The normal pressure on a surface of the plate is $-p_{xx}$, but since v and w vanish over the surface it follows from (2) that $\partial u/\partial x$ also vanishes over the surface, consequently the normal pressure on the plate is simply p .

Thus the difference Δp in the normal pressures on the two surfaces is given by (4a) as 2μ times the algebraic difference of the fields on either side in the electrostatic case; that is to say the normal force acting on unit area of the plate is

$$\Delta p = 8\pi\mu\sigma, \quad \dots \dots \dots (5)$$

where σ is the total charge density (both surfaces included) at the same point on the corresponding conducting lamina.

Fig. 1.



Showing (a) the potential distribution round the mid-section of an earthed, conducting lamina whose length is much greater than its breadth, and (b) the stream lines of the derived viscous flow.

§ 2. A THIN PLATE PERPENDICULAR TO A STREAM.

An electrical system is here considered which consists of a region of uniform potential U disturbed by the presence of an earthed conducting lamina. When the corresponding potential ϕ is inserted in (4), the resulting flow system has zero velocity components at the surfaces of the lamina and a uniform velocity U perpendicular to the plane of the lamina at points very distant from it; that is to say it represents the viscous flow round a thin plate having the same form as the conducting lamina, placed perpendicular to a uniform stream of velocity U . The total normal force R on the plate is obtained by integrating Δp over its area; thus it is given by (5) as $8\pi\mu$ times the total charge on the corresponding conducting lamina. Now this charge is equal to CU where C is the electrical capacity of a conducting lamina of the same shape as the plate, so the normal force is

$$R = 8\pi\mu CU. \quad \dots \dots \dots (6)$$

This equation equally represents the normal force on a thin plate moving in the direction of its normal with a velocity U through a liquid at rest. It is remarkable that this simple result has not been mentioned

before, as the problem of a stream flowing past a perpendicular plate has been considered by Oseen (1927), who used his well known "linearized equations" which give the same results as Stokes' equations when the motion is slow or the viscosity high. Oseen was chiefly concerned with the solution of his equations in the case of a vanishingly small viscosity, but it may readily be shown that the integral equations, by which he formulates his results, reduce to equations (4) in the opposite extreme of very slow motion or very high viscosity.

All the results deduced in the present work apply strictly only to infinitely thin plates. But there is reason to suppose that they will also hold for plates of small thickness with considerable accuracy, at any rate if the edges of the plates are rounded. This can be made clear from the consideration of a simple case: a plate in the form of an oblate spheroid of small thickness. The well-known treatment of viscous flow past an ellipsoid (Lamb 1924 b) yields the following expression for the force on an oblate spheroid whose axis of symmetry is parallel to the undisturbed stream:

$$R = 8\pi\mu bU \left\{ \frac{2e^2 - 1}{e^3} \sin^{-1}e + \frac{\sqrt{1 - e^2}}{e^2} \right\}^{-1}, \quad \dots \quad (7)$$

where $2b$ is the diameter of the spheroid and e its eccentricity. If e approaches unity so that the spheroid degenerates into a circular disk, the force becomes $16\mu bU$: a result also obtainable from (6) by giving C the value $2b/\pi$ appropriate to a circular disk. But if the thickness $2b\sqrt{1 - e^2}$ of the spheroid is made equal to one tenth of its diameter, equation (7) gives a force $16.07\mu bU$ in close agreement with the result for zero thickness.

§ 3. THE CALCULATION OF ELECTRICAL CAPACITIES.

An explicit expression for the capacity of a thin plate appears to be obtainable only when the plate has a circular or elliptic boundary; the capacity of an elliptic plate of major axis a and eccentricity e being $aF(e)$, where $F(e)$ is the complete elliptic integral of the first kind (Rayleigh 1926). Equation (6) thus gives the viscous resistance to an elliptic plate held perpendicular to a stream as

$$R = 8\pi\mu aU/F(e). \quad \dots \quad (6a)$$

There are, however, certain approximate methods by which the capacities of conductors of any shape may be estimated (Maxwell 1892). It is sufficient for the present purpose to state the theorem that the capacity of a plate of any shape must be less than that of a plate whose boundary can be circumscribed about it. For if the boundary of a thin conducting plate be moved outwards, the charge q upon it will redistribute itself with a reduction in its potential energy $\frac{1}{2}q^2/C$; thus C must increase whenever the boundary is moved outwards.

It follows from this theorem that the capacity of a square plate of side $2b$ must lie between $2b/\pi$ and $2\sqrt{2}b/\pi$: the capacities of inscribed and circumscribed circular plates respectively. In the case of a rectangular plate of length $2a$ and breadth $2b$, the inscribed elliptic plate having the greatest capacity has major and minor axes of lengths a and b respectively; so a lower bound to the capacity of the rectangular plate is

$$C_1 = a/F(\sqrt{1-k^2}),$$

where k is the side ratio b/a . Now a circumscribing ellipse of eccentricity e which touches the corners of the rectangle has a capacity

$$C_2 = \frac{a(1+k^2-e^2)^{1/2}}{(1-e^2)^{1/2}F(e)}.$$

TABLE I.

Calculated Capacities of Rectangular Plates.
(Length of plate $2a$, breadth $2b$, capacity C .)

$a/b :$	1	10	10^2	10^3	10^4
$C/a :$	0.768 ± 0.132	0.317 ± 0.047	0.189 ± 0.022	0.133 ± 0.012	0.102 ± 0.008

The minimum value of this expression can be determined graphically for any particular value of k , giving an upper bound to the capacity of a rectangular plate having this side ratio. Table I. gives upper and lower bounds for the capacity of various rectangular plates calculated in this way.

It can be seen from this table, and also proved from the above expressions for C_1 and C_2 , that as the plates becomes longer the bounds become closer; and since $F(\sqrt{1-k^2})$ approaches $\log 4/k$ when k becomes small, the capacity of a long thin plate approximates to the value $a/\log(4a/b)$.

The foregoing calculations of electrical capacity are given to show how equation (6) may be applied, but there appears to be no experimental data at present available on the resistance to slowly moving plates with which a direct comparison may be made.

§ 4. FLOW THROUGH AN ELLIPTIC APERTURE.

Consideration is only given here to the case of an elliptic aperture in a thin plate, although the flow through apertures of other shapes could be treated by approximate methods similar to those detailed above. The potential in the neighbourhood of an elliptic aperture in a charged conducting lamina is most easily obtained by the methods developed

in the theory of gravitation potential, for which reference may be made to Lamb's work (1924 c) in which these methods are applied to problems connected with the flow of *frictionless* fluids. For the solution of the present problem, it is sufficient to note that the expression

$$\phi_1 = Ax \int_{\lambda}^{\infty} \frac{d\lambda}{\lambda^{3/2}(b^2 + \lambda)^{1/2}(c^2 + \lambda)^{1/2}}, \quad \dots \dots \dots (8)$$

is a solution of Laplace's equation, if λ be a function of the coordinates defined by the relation

$$\frac{x^2}{\lambda} + \frac{y^2}{b^2 + \lambda} + \frac{z^2}{c^2 + \lambda} = 1. \quad \dots \dots \dots (8a)$$

Then if the constants b and c be set equal to the major and minor axes of the elliptic aperture respectively, and if the origin of the coordinates be taken at the centre of the aperture with the y axis parallel to b and the z axis parallel to c ; the potential ϕ_1 vanishes over the surface of the lamina and approaches the value $2Ax/r^3$ when the distance r from the origin is very great. Also, the electric field has a component normal to the lamina given by

$$-\frac{\partial \phi_1}{\partial x} = A \int_{\lambda}^{\infty} \left\{ \frac{1}{(b^2 + \lambda)^{3/2}(c^2 + \lambda)^{1/2}} + \frac{1}{(b^2 + \lambda)^{1/2}(c^2 + \lambda)^{3/2}} \right\} \frac{d\lambda}{\lambda^{3/2}}.$$

The value which this component of the field assumes over the aperture is obtained by putting $\lambda=0$ for the lower limit of the integral, in which case the expression reduces to the constant value $2AE(e)/bc^2$, where $e^2 = (b^2 - c^2)/b^2$ and $E(e)$ is the complete elliptic integral of the second kind.

In order to obtain the required solution for the potential distribution ϕ on one side of the plane of the lamina, the following boundary conditions must be fulfilled: (a) the potential must vanish over the surface of the lamina, (b) the field must become uniform and parallel to the x axis at great distances from the aperture, and (c) because of the symmetry of the distribution on either side of the plane of the lamina, the x -component of the field must vanish over the aperture. These requirements are all satisfied by superposing a uniform field

$$X_0 = -2AE(e)/bc^2,$$

parallel to the x axis on the potential distribution given by (8), thus:

$$\phi = \phi_1 - xX_0. \quad \dots \dots \dots (9)$$

The required potential distribution on the other side of the lamina is the mirror image of this, so the charge density on either side of the lamina at great distances from the aperture is $X_0/4\pi$, and the total charge density is

$$\sigma_0 = X_0/2\pi. \quad \dots \dots \dots (10)$$

In order to obtain the solution of the problem of viscous flow, the value for ϕ given by (9) is inserted in equations (4). This gives the result that at great distances from the aperture the flow is radial and has velocity $-2Ax^2/r^4$; so the volume flowing through the aperture in unit time is

$$Q = -\frac{4}{3}\pi A = \frac{2}{3}\pi X_0 bc^2/E(e).$$

But the pressure difference between the sides of the plate at points well removed from the aperture is given by (5) and (10) as

$$\Delta p = 4\mu X_0,$$

and by combining these two equations, the relation between pressure-difference and rate of flow is obtained as

$$\Delta p = \frac{6\mu}{\pi} \frac{E(e)}{bc^2} Q.$$

This result may be put in a simpler form by noting that the perimeter s of the elliptic aperture is equal to $4bE(e)$, so that

$$\Delta p = \frac{3}{2}\pi\mu \frac{s}{S^2} Q, \quad \dots \dots \dots (11)$$

where S is the area of the aperture.

In the particular case of a *circular aperture* of radius c , the above relation becomes

$$\Delta p = 3\mu Q/c^3. \quad \dots \dots \dots (11a)$$

The problem of the flow of a viscous fluid through a circular aperture has already been solved by Sampson (1891), as a particular case of the flow of a fluid bounded by a hyperboloid of one sheet. Sampson obtained an expression similar to (11a), but differing by a factor of 4 on the right-hand side as a result of trivial errors made in setting down his expressions for the stream function and total flux.

The relation between pressure difference and rate of flow has been determined experimentally by different investigators, who make no reference to Sampson's theoretical work. Thus Bond (1922) observed the rate of flow of glycerine-water mixtures through a hole 0.1469 cm. in diameter drilled in a plate 0.0075 cm. thick. He found the pressure difference could be expressed by the formula $(16k/\pi)(Q\mu/c^3)$ with $k = 0.631 \pm 0.01$, which is equivalent to (11a) with the constant equal to 3.21 ± 0.05 instead of the theoretical value of 3. He attempted to correct his result so as to find k for an infinitely thin plate (obtaining a value of 0.580 ± 0.01 which is in good agreement with the present theory), but the correction he employed requires justification. Johansen (1930) has studied the flow of castor oil through circular orifices situated within a pipe, the orifices having sharp edges being bevelled at 45° on the low pressure side. The smallest orifice used by him had a diameter of 0.090

times the diameter of the pipe, and it is evident from his results with different orifice diameters that the finite width of the pipe does affect the result in this case. By measurement of his graphical presentation of the results, a figure of 2.99 is obtained to correspond with the theoretical figure 3 in equation (11a); the accuracy of his observations being probably about the same as that of Bond's. In the results of both writers, a radical departure from the proportionality between p and Q is observable at high rates of flow, this departure becoming noticeable in the observations when the dimensionless quantity $\frac{Q\rho}{c\mu}$ exceeds 10 (ρ being the density of the liquid).

§ 5. FLOW THROUGH A SLIT.

It is evident from dimensional considerations that the relation between the rate of flow per unit length Q' and pressure difference, in the case of a long slit with parallel sides, will be of the form

$$Q' = K \Delta p t^3 / \mu,$$

where t is the width of the slit, and K is some constant. From this it follows that the total rate of flow Q through a long elliptic aperture is $2bK \Delta p / \mu$ times the mean square breadth. The latter is $\frac{8}{3} c^2$, so that

$$Q = \frac{16}{3} \frac{K \Delta p b c^2}{\mu} = \frac{64}{3\pi^2} \frac{K \Delta p S^2}{\mu s}.$$

Comparing this result with the general expression for an elliptic aperture (11), it is seen that K is equal to $\pi/32$. So for the long slit with parallel sides we have

$$\Delta p = \frac{32}{\pi} \frac{\mu Q'}{t^3}. \quad \dots \dots \dots (12)$$

§ 6. ROTATING PLATES.

Suppose an earthed conducting lamina be placed in a uniform electric field of strength Ω , so that its plane contains the direction of the field. Then if the origin of coordinates be taken at some point in the plane of zero potential and the y axis parallel to the field (fig. 2), the undisturbed potential has the form $\phi_0 = -\Omega y$. Now if the actual potential ϕ (made up of ϕ_0 and the disturbing potential produced by the earthed lamina), be inserted in equations (4) a flow system results which, over the surface of the lamina has zero velocity, but at great distances from the lamina has components

$$u = -\Omega y, \quad v = \Omega x, \quad w = 0.$$

Thus the flow system obtained is that of a fluid rotating with angular velocity Ω about the z axis round a fixed plate having the same form as the conducting lamina.

The plate can be held fixed in this position by a suitable torque (G) applied about the z axis, together with a force (R) acting at some point on this axis, and parallel with the x axis. From (5), the magnitudes of these quantities are obtained as

$$R = 8\pi\mu Q,$$

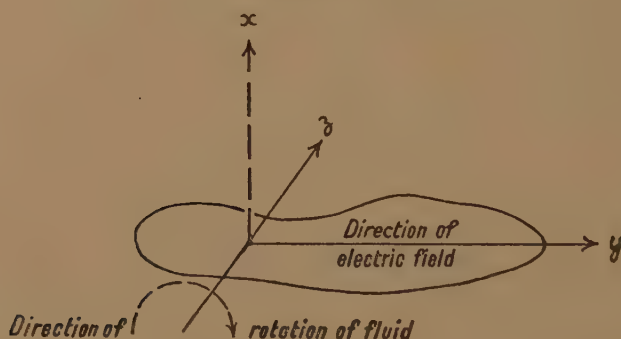
$$G = 8\pi \int y\sigma dS,$$

where Q is the total charge on the corresponding earthed lamina, σ the charge density and dS a surface element. The latter equation may be written

$$G = 8\pi\mu M\Omega, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where M represents the moment about the z axis of the charge distribution induced on the earthed lamina by an electric field of unit strength acting in the y direction and having zero potential over the plane $y=0$.

Fig. 2.



If a plate is rotated with angular velocity Ω in a stationary fluid of infinite extent, the torque is also given by equation (13). The case of an elliptic plate rotating about its minor axis is considered here as a simple example of the application of this equation. The constant M is in this case equal to the electric dipole moment acquired by an earthed elliptic lamina, with major and minor axes b and c respectively parallel to the y and z axes, when placed in an electric field of unit strength parallel to the y axis. Now the solution of Laplace's equation represented by

$$\phi = x \int_0^\lambda \frac{d\lambda}{\lambda^{1/2}(b^2 + \lambda)^{3/2}(c^2 + \lambda)^{1/2}}, \quad . \quad . \quad . \quad . \quad (14)$$

where λ is defined as before by (8a), vanishes over the surface of the lamina. Furthermore, at a great distance r from the centre of the lamina it approaches the value

$$x \int_0^\infty \frac{d\lambda}{\lambda^{1/2}(b^2 + \lambda)^{3/2}(c^2 + \lambda)^{1/2}} - \frac{2}{3} \frac{x}{r^3}.$$

This expression represents the potential due to a uniform field equal in magnitude to the integral, disturbed by an electric dipole of moment $2/3$. So in a field of unit strength the lamina will acquire a charge distribution whose dipole moment M is equal to $2/3$ of the reciprocal of this integral, that is to say

$$\frac{2}{3M} = \int_0^\infty \frac{d\lambda}{\lambda^{1/2}(b^2 + \lambda)^{3/2}(c^2 + \lambda)^{1/2}} = \frac{2}{b^3 e^2} \{F(e) - E(e)\}, \quad \dots \quad (15)$$

where $e^2 = (b^2 - c^2)/b^2$ and $F(e)$ and $E(e)$ are the complete elliptic integrals of the first and second kind respectively.

The ellipse degenerates into a circle of radius c as e approaches zero, and M is then given by (15) as $4c^3/3\pi$. On inserting this result in (13) it is seen that the torque on a circular plate of radius c rotating about a diameter in a viscous liquid is

$$G = \frac{32}{3} \mu c^3 \Omega.$$

Similarly, the result for a very long elliptic plate rotating about its minor axis is obtained by making e very large, in which case

$$G = \frac{8}{3} \frac{\pi \mu b^3 \Omega}{\log(4b/c) - 1}.$$

In the case of an elliptic plate rotating about a tangent at one end of its major axis, the charge moment M to be inserted in (13) is made up of the induced dipole moment in unit field, together with the moment of the charge distribution acquired in a region of uniform potential b , that is to say,

$$M = \frac{b^3 e^2/3}{F(e) - E(e)} + \frac{b^3}{F(e)}. \quad \dots \quad (16)$$

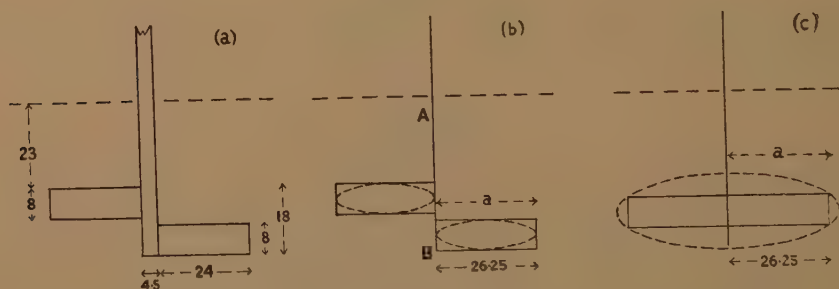
The moment M of the charge distribution may be calculated approximately for other boundary shapes than the ellipse, by similar methods to those described in § 3 for the calculation of electrical capacities. Thus the value of M for a particular lamina must be less than the charge moment of any other lamina lying in the same plane whose boundary circumscribes it, and greater than that of any other lamina lying in the same plane whose boundary is inscribed within it, a result which can be proved as follows: the potential energy of an earthed lamina in an electric field is negative and in magnitude equal to half the product of M and the square of the field strength. If the field be maintained constant and the boundary of the lamina be moved outward, the charge upon it will redistribute itself under the influence of electrical forces, that is to say, with a reduction in the potential energy of the system. Consequently the value of M is increased by any outward motion of the boundary, and hence it must be decreased by any inward motion. An application of these results is given in the next section.

§ 7. A PADDLE VISCOMETER.

A modification of the Stormer Viscometer, described by Geddes and Dawson (1942), consists of a paddle (fig. 3 *a*) rotating in a vessel containing the liquid under test. The drive is provided by a falling weight attached to a string passing over a pulley and round a drum of one inch diameter. A gear wheel with 275 teeth fixed to this drum engages with another (having 25 teeth) attached to the paddle spindle. A revolution counter is also attached to the paddle spindle. The viscosity of the liquid is generally assumed to be inversely proportional to the number of revolutions per second made by the paddle with a given driving weight.

Such instruments are used industrially to test the flow properties of paints and oils. Although these liquids are not always Newtonian, and their "apparent viscosities" are customarily read on a purely arbitrary scale, it has been shown by Geddes and Dawson that their

Fig. 3.



The paddle of Geddes and Dawson's viscometer (*a*), and the forms used in calculation (*b* & *c*). The dimensions shown are in millimetres.

instrument may be calibrated by means of a liquid of standard viscosity so as to give a true measure of viscosity in c.g.s. units for Newtonian liquids, provided certain necessary corrections are made in calculating the results.

It is of interest to use the mathematical method described in § 6 above to determine the relation between the viscosity of the liquid and the rate of rotation of the paddle under a given driving weight, in the case of slow rotation, to show how an apparently complicated hydrodynamic problem is soluble in a simple although approximate manner by this method. The solution of this theoretical problem has, moreover, a practical bearing on the use of the instrument. For the relation between rate of rotation and driving weight is not strictly linear, even for Newtonian liquids, within the range of driving weights commonly used; and Geddes and Dawson point out that this implies that the rate of rotation is a function of the density of the liquid as well as of its viscosity, so that "viscosity" as commonly measured by this instrument is in reality a hybrid of density and viscosity. Now at very slow rates of rotation, the mathematical expression for the torque on the paddle would consist of a single term

proportional to the rate of rotation and the viscosity; while at somewhat higher speeds the departure from linear proportionality would be accounted for by an additional term proportional to the square of the rate of rotation and depending on the density. Geddes and Dawson, working on the assumption that their observations lie in this low-speed range, have fitted a quadratic relation to their observations; and from the constants of this empirical relation have deduced the correction factor to be applied to observations made with this viscometer in order that the effect of density may be eliminated and a true measure of viscosity obtained. For the justification of this procedure it is necessary to show that the observations really do lie near to the low-speed region of strict proportionality between torque and rate of rotation. An approximate theoretical calculation for slow rates of rotation will demonstrate whether this is the case, for if the observations really lie in a speed range far removed from the region of strict proportionality, the observed torques will be found to be much in excess of their calculated values.

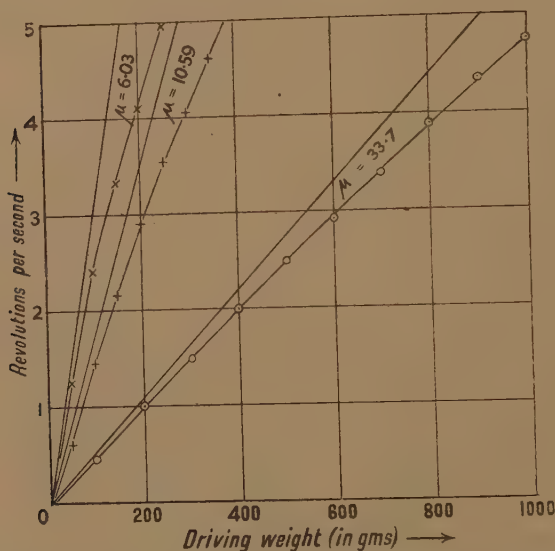
It is not difficult to show theoretically that, in making the calculations, it is permissible to neglect the effect of the walls of the containing vessel and the free surface of the liquid when the normal pint-size container is used, but the proof is unnecessary here, since Geddes and Dawson have shown that the observations are unaffected by the substitution of a quart-size container, although the use of much smaller containers affects the observations appreciably. The central spindle has a negligible effect upon the resistance to rotation, since its radius is small compared with the length of the blades, and the resistance to a rotating body is conditioned much more by its shape at points far removed from the axis of rotation than by its shape in the neighbourhood of the axis. So for the purpose of calculation the shape of the paddle may be taken as that shown by the full lines in fig. 3*b*.

The torque on the paddle being then given by (13), it remains to calculate the charge moment M for a lamina of this form in a field of unit strength lying in its plane and perpendicular to the axis AB . It should first be noted that if the two rectangles could be slid apart along the line AB , work would have to be done in overcoming the mutual attraction of their charge distributions, the potential energy of the system being thereby increased and the value of M consequently decreased. Thus the value of M for the lamina shown is greater than the sum of the charge moments for the two rectangular parts separated by an infinite distance along AB . The charge moment of the latter system is greater again than that of two elliptic laminæ inscribed within the rectangles (as shown by the broken lines in fig. 3*b*), also removed to an infinite distance apart along AB , by the theorem given at the end of § 6. The axial ratio of the ellipses is 3.28, and on inserting this figure in equation (16) the sum of the charge moments of the two elliptic laminæ is obtained as $1.14a^3$, where a represents the distance of the extreme edges of the rectangles from the axis. This quantity is therefore a lower bound to the value of M .

It follows by a similar argument that the value of M is less than the charge moment of the single rectangle shown by the full line in fig. 3c, which again is less than that of any elliptic lamina whose boundary circumscribes it. The charge moments of various circumscribing elliptic laminae of different eccentricities can be calculated from (15), and it can then be found by graphical interpolation that of all such laminae, that with an eccentricity of 0.94 (shown by the broken line in fig. 3c) has the lowest charge moment, that is $2.25a^3$. The latter figure therefore gives an upper bound to M . Combining this with the lower bound found previously, it is seen from (13) that the torque can be written as

$$G = (1.70 \pm 0.55) 8\pi\mu a^3 \Omega.$$

Fig. 4.



Calculated and observed results for Geddes and Dawson's viscometer. (The viscosities shown against the curves are in poises.)

The bounds within which the true value of G must lie, as given by this expression, are very wide; but the true value will certainly lie closer to the mean than these bounds imply, since it is clear from the method of deduction that one bound must lie considerably above the true value and the other considerably below.

To compare this calculated result with the experimental observations, it is convenient to convert the torque into driving weight (expressed in grams), by multiplying it by the gear ratio and dividing by the product of g and the radius of the drum; and to express angular velocity as revolutions per second. The resulting relations between rate of rotation and driving weight, calculated for liquids of various viscosities, are shown by the straight lines in fig. 4, together with experimental observations of Geddes and Dawson. At the lower speeds of rotation the experimental

results lie quite close to the theoretical straight lines which, of course, are calculated on the assumption of low speeds. It may also be mentioned that the theoretical straight lines are very close to straight lines obtained by Geddes and Dawson when they apply their empirical correction to the rate of rotation before plotting their results. This justifies the method of correction suggested by those writers.

The author wishes to thank Prof. T. H. Havelock, F.R.S., for his kindness in reading a draft of this paper and suggesting certain improvements.

REFERENCES.

- BOND, W. N., 1921, *Proc. Phys. Soc.*, **33**, 225; 1922, *Ibid.*, **34**, 139.
 GEDDES, J. A., and DAWSON, D. H., 1942, *Ind. Eng. Chem.*, **34**, 163.
 JOHANSEN, F. C. 1930, *Proc. Roy. Soc. A*, **126**, 231.
 LAMB, H., 1924 a, *Hydrodynamics*, 5th Ed., p. 547 (Cambridge); 1924 b, *Ibid.*, p. 571; 1924 c, *Ibid.*, p. 139.
 MAXWELL, J. C., 1892, *Treatise on Electricity and Magnetism*, 3rd Ed., **1**, 148.
 OSEEN, C. W., 1927, *Hydrodynamik*, p. 214 (Leipzig).
 RAYLEIGH, Lord, 1926, *Theory of Sound*, 2nd Ed., **2**, 176.
 SAMPSON, R. A., 1891, *Phil. Trans. A*, **182**, 449.

XXXII. Tables of the Function $\int_0^x \frac{-\log|1-y|}{y} dy$, with an Account of Some Properties of this and Related Functions.

By K. MITCHELL, Ph.D., A.F.R.Ae.S.,
 King's College, Newcastle-upon-Tyne*.

[Received April 9, 1948.]

INTRODUCTION.

INTEGRALS of the form

$$\int \frac{\log(ax+b)}{cx+d} dx, \quad (1.1)$$

and of forms reducible thereto, occur in certain physical problems. Substituting $y=a(cx+d)/(ad-bc)$, we find

$$\int \frac{\log(ax+b)}{cx+d} dx = \frac{1}{c} \log \left| b - \frac{ad}{c} \right| \log \left| \frac{a(cx+d)}{ad-bc} \right| - \frac{1}{c} f \left\{ \frac{a(cx+d)}{ad-bc} \right\}, \quad . . (1.2)$$

where

$$f(x) = \int_0^x \frac{-\log|1-y|}{y} dy. \quad (1.3)$$

All integrals of the form (1.1) can therefore be expressed in terms of a function of a single parameter, a possible standard form of which is given

* Communicated by the Author.

in (1.3). If we extend the definition of the function to complex values, too, it will be possible, with its aid, to evaluate the double integral of any rational function.

Some properties of an integral of this type, the function

$$Rl(x) = \int_1^x \frac{\log \xi \, d\xi}{\xi - 1} = -f(1-x), \quad . \quad . \quad . \quad . \quad (1.4)$$

have been given (Powell 1943) together with a table of the function over the range $x=0.00$ (0.01) 2.00 (0.02) 6.00 : 7D. An earlier table (Newman 1892) gives the function $Rl(1 \pm x)$ for $x=0.00$ (0.01) 0.50 : 12D, and a comparison of these two tables over their common range (Fletcher 1944) reveals discrepancies which are traced to errors in Newman's table. Powell's table, over the common range, contains only rounding errors in the last decimal place. Another table (Spence 1809) gives $Rl(x)$, $x=1$ (1) 100 : 9D; yet another (Kummer 1840) gives a function akin to $f(x)$ for $x=-11$ (1) 10 : 11D. These tables offer only isolated points of comparison.

The author's attention was drawn to the function by a physical problem in 1940, and tables of $f(x)$ covering the ranges $x=-1.00$ (0.01) 1.00 : 9D; and $x=0.000$ (0.001) 0.500 : 9D, were calculated. These tables, which supplement those of Powell and Newman, are given at the end of this paper.

The standard function was taken in the form (1.3) and the author has preferred to retain that form, rather than to adopt any of the earlier definitions, which differ among themselves. His preference is founded on the relationship the integral bears to the hypergeometric function, and to the ζ -function of Riemann (below, §§ 2 and 10).

The paper also contains a short account of the properties of the function $f(x)$, defined as a function of a complex variable. Many of the results given were known to Spence, Kummer and others (Hill 1828 and 1830) for real values of the variable. Kummer defines his function for complex values, and splits it into real and imaginary parts, but his consideration of the complex domain goes no further than this.

Comparison of the author's table with that of Powell reveals one error in the latter :

at $x=0.01$ (Powell's variable), for 1.5886 274 read 1.5886 254 :

there are also several rounding errors in the last decimal place, which have not been separately listed. Errors in Newman's table are listed by Fletcher.

DEFINITION AND SIMPLE PROPERTIES.

2. For complex z , not real and greater than unity, we define

$$f(z) = \int_0^z \frac{-\log(1-\zeta)}{\zeta} d\zeta, \quad . \quad . \quad . \quad . \quad (2.1)$$

where $\log(1-\zeta) = \log|1-\zeta| + i \arg(1-\zeta)$, $|\arg(1-\zeta)| < \pi$, and the path of integration is the straight line from the origin of the complex

plane to the point z . This defines a function regular everywhere in a complex plane cut from $+1$ to infinity along the real axis. To agree with (1.3) the function must be defined, for real values of z greater than unity, as the limiting value of the arithmetic mean of values immediately above and below the real axis.

For values of z of modulus not greater than unity the function possesses the absolutely and uniformly convergent expansion

[illegible]

from which we deduce the special values

$$\left. \begin{aligned} f(1) &= \frac{1}{6} \pi^2 \\ f(-1) &= -\frac{1}{12} \pi^2 \end{aligned} \right\} \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (2.3)$$

We note also that (2.2) is a degenerate hypergeometric series,

$$\sum_{n=1}^{\infty} \frac{z^n}{n!} = \lim_{\epsilon \rightarrow 0} \{ {}_2F_1(\epsilon, \epsilon; 1; z) - 1 \}. \quad (2.4)$$

This suggests the existence of the properties which we shall next establish.

3. Suppose z is not real. Then by deforming the path of integration into segments of straight lines joining the origin to the point 1, and 1 to z , and the part of the circumference of the circle $|z-1|=\epsilon$ joining these lines and not crossing the cut in the real axis, and proceeding to the limit as ϵ tends to zero, we find

$$f(z) = \int_0^1 \frac{-\log(1-\zeta)}{\zeta} d\zeta + \int_1^z \frac{-\log(1-\zeta)}{\zeta} d\zeta. \quad (3.1)$$

The first integral has the value $\pi^2/6$ by (2.3). Integrating the second by parts (taking $\log \zeta = \log |\zeta| + i \arg \zeta$, $|\arg \zeta| < \pi$),

$$f(z) = \frac{1}{6}\pi^2 - \log z \log(1-z) + \int_1^z \frac{-\log \zeta}{1-\zeta} d\zeta. \quad (3.2)$$

Taking $1 - \zeta$ as a new variable of integration, this leads to

$$f(z)+f(1-z)+\log z \log (1-z)=\frac{1}{6} \pi^2 . \quad . \quad . \quad . \quad (3.3)$$

The restriction to non-real values of z may be removed by continuity arguments, and for real values of x we find

$$f(x) + f(1-x) + \log |x| \log |1-x| = \frac{1}{6}\pi^2. \quad (3.4)$$

This gives us two further special values. Putting $x=\frac{1}{2}$,

$$f(\frac{1}{2}) = \frac{1}{12} \pi^2 - \frac{1}{2} \log^2 2, \quad . \quad . \quad . \quad . \quad . \quad (3.5)$$

and putting $x = -1$ and using (2.3),

$$f(2) = \frac{1}{4}\pi^2. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (3.6)$$

4. Suppose again that z is not real. Then on proceeding as above

$$\begin{aligned} f(1/z) &= f(1) + \int_1^{1/z} \frac{-\log(1-\zeta)}{\zeta} d\zeta, \\ &= f(1) - \int_1^z \frac{-\log(1-1/\eta)}{\eta} d\eta, \end{aligned}$$

on making the transformation $\eta = 1/\zeta$. The path of integration of the transformed expression is curved, but may be replaced by the straight line from 1 to z , since the integrand is analytic within the region between these paths. Now we have

$$\log(1-1/\eta) = \log(1-\eta) - \log(-\eta),$$

where each of the logarithms has its principal value. Hence

$$\begin{aligned} f(1/z) &= f(1) - \int_1^z \frac{\log(-\eta) - \log(1-\eta)}{\eta} d\eta, \\ &= 2f(1) - f(z) - \left[\frac{1}{2} \log^2(-\eta) \right]_1^z, \end{aligned}$$

which may be written

$$f(1/z) + f(z) + \frac{1}{2} \log^2(-z) = -\frac{\pi^2}{6}. \quad . \quad . \quad . \quad (4.1)$$

When z is real and negative a limiting argument leads without difficulty to

$$f(1/x) + f(x) + \frac{1}{2} \log^2(-x) = -\frac{\pi^2}{6}, \quad x < 0. \quad . \quad . \quad . \quad (4.2)$$

When x is real and positive it is necessary to take the arithmetic mean above and below the cut. We then find

$$f(1/x) + f(x) + \frac{1}{2} \log^2 x = \frac{\pi^2}{3}, \quad x > 0. \quad . \quad . \quad . \quad (4.3)$$

5. By combining the two relations so far obtained, others can be derived which connect functions having the arguments z , $1-z$, $1/z$, $1/(1-z)$, $(z-1)/z$, $z/(z-1)$. In complex form, these additional relations are

$$f(z) - f\left(\frac{z-1}{z}\right) + \frac{1}{2} \log^2(-z) + \log z \log\left(\frac{z-1}{z}\right) = -\frac{\pi^2}{3}, \quad . \quad . \quad . \quad (5.1)$$

$$\begin{aligned} f(z) + f\left(\frac{z}{z-1}\right) + \frac{1}{2} \log^2\left(\frac{1-z}{z}\right) + \log z \log\left(\frac{z-1}{z}\right) + \frac{1}{2} \log^2(-z) &= -\frac{\pi^2}{2}, \\ &. \quad . \quad . \quad . \quad . \quad . \quad (5.2) \end{aligned}$$

$$f(z) - f\left(\frac{1}{1-z}\right) + \log z \log(1-z) - \frac{1}{2} \log^2(z-1) = \frac{\pi^2}{3}, \quad . \quad . \quad . \quad (5.3)$$

where each logarithm has its principal value. In writing the real forms,

some simplification of the logarithmic terms can be made. We find

$$\left. \begin{aligned} f(x) - f\left(\frac{x-1}{x}\right) + \frac{1}{2} \log |x| \log \frac{(x-1)^2}{|x|} &= \frac{\pi^2}{6}, & x > 0 \\ &= -\frac{\pi^2}{3}, & x < 0 \end{aligned} \right\}, \quad (5.4)$$

$$\left. \begin{aligned} f(x) + f\left(\frac{x}{x-1}\right) + \frac{1}{2} \log^2 |x-1| &= \frac{\pi^2}{2}, & x > 1 \\ &= 0, & x < 1 \end{aligned} \right\}, \quad (5.5)$$

$$\left. \begin{aligned} f(x) - f\left(\frac{1}{1-x}\right) + \frac{1}{2} \log |1-x| \log \frac{x^2}{|1-x|} &= \frac{\pi^2}{3}, & x > 1 \\ &= -\frac{\pi^2}{6}, & x < 1 \end{aligned} \right\}. \quad (5.6)$$

These formulæ indicate that, if values for real values of x are required, it is necessary to calculate the function $f(x)$ only for the range $0 \leq x \leq \frac{1}{2}$. Values for other ranges can be obtained directly from these and tables of natural logarithms, most easily by the formulæ as indicated below:—

Range of x	$x \leq -1$	$-1 \leq x \leq 0$	$\frac{1}{2} \leq x \leq 1$	$1 \leq x \leq 2$	$2 \leq x$
Formula	(5.6)	(5.5)	(3.3)	(5.4)	(4.3)

6. The above properties are suggested by the hypergeometric relationship. Another property is obtained as follows. We have

$$f(z^2) = \int_0^{z^2} \frac{-\log(1-\zeta)}{\zeta} d\zeta.$$

Changing the variable of integration to η , where $\eta^2 = \zeta$, and using the fact that

$$\log(1-\eta^2) = \log(1-\eta) + \log(1+\eta),$$

each logarithm having its principal value,

$$f(z^2) = 2 \int_0^z \frac{-\log(1-\eta)}{\eta} d\eta + 2 \int_0^z \frac{-\log(1+\eta)}{\eta} d\eta.$$

Changing the sign of the variable of integration in the second integral,

$$f(z) + f(-z) = \frac{1}{2} f(z^2). \quad . \quad . \quad . \quad . \quad . \quad . \quad (6.1)$$

The obvious generalization,

$$\frac{1}{n} f(z^n) = \sum_{r=1}^n f\{z \exp(2\pi r i/n)\}, \quad . \quad . \quad . \quad . \quad . \quad (6.2)$$

for positive integral values of n , is similarly proved.

TABULATION OF THE FUNCTION.

7. The power series (2.2) is reasonably convenient for the calculation of $f(z)$ for real values of z in the range 0 to $\frac{1}{2}$, but a more rapidly convergent series is easily obtained. We have, when $|z| \leq 1$,

$$\left. \begin{aligned} \sum_{n=1}^{\infty} \frac{z^n}{n^2} &= f(z), & \sum_{n=1}^{\infty} \frac{z^n}{(n+1)^2} &= -1 + \frac{1}{z} f(z) \\ \sum_{n=1}^{\infty} \frac{z^n}{n(n+1)} &= \frac{1-z}{z} \log(1-z) + 1 \end{aligned} \right\}.$$

We hence form the series

$$g(z) = \sum_{n=1}^{\infty} \frac{z^n}{n^2(n+1)^2}, \quad \dots \dots \dots (7.1)$$

and find

$$f(z) = \frac{z\{3+g(z)\} + 2(1-z) \log(1-z)}{1+z}. \quad \dots \dots \dots (7.2)$$

The function $g(z)$ was evaluated for $z=0.00$ (0.01) 0.50 : 1.0D, and with this and values of the natural logarithm to 10D a basic table of $f(z)$ was constructed, also to 10D. Differencing of this table suggested* that the values obtained were correct to within one or possibly two units in the tenth place (the best that could be expected in view of rounding errors in the logarithms) and the table was subsequently subtabulated to tenths by the end-figure process. The resulting table is given as Table 2 herewith, rounded to 9D.

Table I. contains, also to 9D, values of the function for the range $z=-1.00$ (0.01) 1.00. Those for negative values of z , from -0.40 to -1.00 , were obtained to 10D by (5.5), calculating by interpolating in the ten-figure version of Table II. (The method of iterative linear interpolation was found most convenient, on account of the largeness of the second difference in this table in comparison with the third.) For the range $z=0$ to -0.40 the series (7.2) was used, and for the range $z=0.50$ to 1.00, the formula (3.3) was used. Differencing the tables again suggested errors of only one or two units in the tenth place, except near $z=1.00$, where, because of the infinity of the derivative, the differences are too large to afford a check. Differencing was thought to be a sufficient check as far as $z=0.90$ (though the formula (6.1) was also applied at all the points $z=0, 0.1, 0.2, \dots, 1.0$), and all the values of $f(z)$ beyond this value of the argument were checked by calculating $f(z^2)$ by (6.1), and directly by (3.3), interpolating in Table II. to obtain $f(1-z)$.

* This was subsequently confirmed by comparison with Newman's table.

TABLE I.

Values of $\zeta(1, 2|z) = \int_0^z \frac{-\log(1-y)}{y} dy$, $z = -1.00$ (0.01) 1.00 : 9D.

z	$\zeta(1, 2 z) = -\text{Rl}(1-z)$	z	$\zeta(1, 2 z) = -\text{Rl}(1-z)$
-1.0 0	-0.8 2246 7033	-0.5 0	-0.4 4841 4207
-0.9 9	1552 5881	-0.4 9	4029 0434
8	0856 5277	8	3213 7541
7	-0.8 0158 5083	7	2395 5256
6	-0.7 9458 5157	6	1574 3304
5	8756 5359	5	-0.4 0750 1404
4	8052 5544	4	-0.3 9922 9272
3	7346 5566	3	9092 6620
2	6638 5279	2	8259 3153
1	5928 4534	1	7422 8573
-0.9 0	-0.7 5216 3179	-0.4 0	-0.3 6583 2578
-0.8 9	4502 1063	-0.3 9	5740 4858
8	3785 8030	8	4894 5101
7	3067 3924	7	4045 2988
6	2346 8588	6	3192 8195
5	1624 1859	5	2337 0394
4	0899 3577	4	1477 9249
3	-0.7 0172 3576	3	-0.3 0615 4420
2	-0.6 9443 1691	2	-0.2 9749 5560
1	8711 7752	1	8880 2318
-0.8 0	-0.6 7978 1588	-0.3 0	-0.2 8007 4334
-0.7 9	7242 3027	-0.2 9	7131 1244
8	6504 1893	8	6251 2677
7	5763 8009	7	5367 8254
6	5021 1195	6	4480 7592
5	4276 1269	5	3590 0298
4	3528 8045	4	2695 5973
3	2779 1338	3	1797 4213
2	2027 0957	2	-0.2 0895 4602
1	1272 6710	1	-0.1 9989 6720
-0.7 0	-0.6 0515 8402	-0.2 0	-0.1 9080 0138
-0.6 9	-0.5 9756 5836	-0.1 9	8166 4417
8	8994 8812	8	7248 9113
7	8230 7128	7	6327 3771
6	7464 0576	6	5401 7928
5	6694 8948	5	4472 1112
4	5923 2034	4	3538 2840
3	5148 9619	3	2600 2623
2	4372 1484	2	1657 9959
1	3592 7411	1	-0.1 0711 4337
-0.6 0	-0.5 2810 7174	-0.1 0	-0.0 9760 5235
-0.5 9	2026 0547	-0.0 9	8805 2122
8	1238 7300	8	7845 4453
7	-0.5 0448 7198	7	6881 1675
6	-0.4 9656 0005	6	5912 3220
5	8860 5481	5	4938 8510
4	8062 3380	4	3960 6955
3	7261 3456	3	2977 7950
2	6457 5457	2	1990 0879
1	5650 9127	1	0997 5110
-0.5 0	-0.4 4841 4207	-0.0 0	-0.0 0000 0000

TABLE I. (*cont.*)

z	$\zeta(1, 2 z) = -Rl(1-z)$	z	$\zeta(1, 2 z) = -Rl(1-z)$
0.0 0	0.0 0000 0000	0.5 0	0.5 8224 0526
1	1002 5112	1	0.5 9616 5361
2	2010 0899	2	0.6 1021 6108
3	3022 8052	3	2439 6071
4	4040 7275	4	3870 8705
5	5063 9292	5	5315 7631
6	6092 4842	6	6774 6644
7	7126 4682	7	8247 9725
8	8165 9588	8	0.6 9736 1058
0.0 9	0.0 9211 0353	0.5 9	0.7 1239 5042
0.1 0	0.1 0261 7791	0.6 0	0.7 2758 6308
1	1318 2737	1	4293 9737
2	2380 6046	2	5846 0483
3	3448 8595	3	7415 3992
4	4523 1283	4	0.7 9002 6024
5	5603 5034	5	0.8 0608 2689
6	6690 0794	6	2233 0471
7	7782 9536	7	3877 6261
8	8882 2258	8	5542 7404
0.1 9	0.1 9987 9987	0.6 9	7229 1733
0.2 0	0.2 1100 3775	0.7 0	0.8 8937 7624
1	2219 4708	1	0.9 0669 4053
2	3345 3898	2	2425 0654
3	4478 2492	3	4205 7798
4	5618 1667	4	6012 6675
5	6765 2639	5	7846 9393
6	7919 6656	6	0.9 9709 9088
7	0.2 9081 5005	7	1.0 1603 0062
8	0.3 0250 9012	8	3527 7934
0.2 9	1428 0044	0.7 9	5485 9830
0.3 0	0.3 2612 9510	0.8 0	1.0 7479 4600
1	3805 8866	1	1.0 9510 3088
2	5006 9611	2	1.1 1580 8451
3	6216 3296	3	3693 6560
4	7434 1521	4	5851 6487
5	8660 5941	5	1.1 8058 1124
6	0.3 9895 8267	6	1.2 0316 7961
7	0.4 1140 0269	7	2632 0101
8	2393 3778	8	5008 7584
0.3 9	3656 0692	0.8 9	7452 9160
0.4 0	0.4 4928 2974	0.9 0	1.2 9971 4723
1	6210 2664	1	1.3 2572 8728
2	7502 1875	2	5267 5161
3	0.4 8804 2799	3	1.3 8068 5041
4	0.5 0116 7714	4	1.4 0992 8300
5	1439 8989	5	4063 3797
6	2773 9085	6	1.4 7312 5860
7	4119 0562	7	1.5 0789 9041
8	5475 6089	8	4579 9712
0.4 9	6843 8444	0.9 9	1.5 8862 5448
0.5 0	0.5 8224 0526	1.0 0	1.6 4493 4067

TABLE II.

Values of $\zeta(1, 2|z) = \int_0^z \frac{-\log|1-y|}{y} dy$, $z=0.000$ (0.001) 0.500 : 9D.

z	$\zeta(1, 2 z) = -\text{Rl}(1-z)$	z	$\zeta(1, 2 z) = -\text{Rl}(1-z)$
0.00 0	0.00 0000 000	0.05 0	0.05 6039 292
1	1000 250	1	1665 426
2	2001 001	2	2692 096
3	3002 253	3	3719 302
4	4004 007	4	4747 046
5	5006 264	5	5775 329
6	6009 024	6	6804 150
7	7012 288	7	7833 512
8	8016 057	8	8863 414
0.00 9	0.00 9020 331	0.05 9	0.05 9893 857
0.01 0	0.01 0025 112	0.06 0	0.06 0924 842
1	1030 399	1	1956 371
2	2036 193	2	2988 443
3	3042 496	3	4021 059
4	4049 307	4	5054 221
5	5056 628	5	6087 928
6	6064 459	6	7122 182
7	7072 801	7	8156 984
8	8081 655	8	0.06 9192 334
0.01 9	0.01 9091 020	0.06 9	0.07 0228 234
0.02 0	0.02 0100 899	0.07 0	0.07 1264 682
1	1111 291	1	2301 682
2	2122 198	2	3339 233
3	3133 620	3	4377 336
4	4145 557	4	5415 993
5	5158 011	5	6455 203
6	6170 982	6	7494 967
7	7184 471	7	8535 287
8	8198 478	8	0.07 9576 164
0.02 9	0.02 9213 005	0.07 9	0.08 0617 597
0.03 0	0.03 0228 052	0.08 0	0.08 1659 588
1	1243 619	1	2702 137
2	2259 708	2	3745 246
3	3276 319	3	4788 915
4	4293 452	4	5833 145
5	5311 110	5	6877 937
6	6329 291	6	7923 292
7	7347 998	7	0.08 8969 210
8	8367 231	8	0.09 0015 692
0.03 9	0.03 9386 989	0.08 9	1062 739
0.04 0	0.04 0407 275	0.09 0	0.09 2110 353
1	1428 089	1	3158 533
2	2449 432	2	4207 280
3	3471 304	3	5256 596
4	4493 706	4	6306 481
5	5516 639	5	7356 936
6	6540 103	6	8407 962
7	7564 100	7	0.09 9459 560
8	8588 630	8	0.10 0511 730
0.04 9	0.04 9613 694	0.09 9	1564 473
0.05 0	0.05 0639 292	0.10 0	0.10 2617 791

TABLE II. (cont.)

z	$\zeta(1, 2 z) = -\text{Rl}(1-z)$	z	$\zeta(1, 2 z) = -\text{Rl}(1-z)$
0.10 0	0.10 2617 791	0.15 0	0.15 6035 034
1	3671 684	1	7118 804
2	4726 153	2	8203 194
3	5781 198	3	0.15 9288 207
4	6836 821	4	0.16 0373 843
5	7893 023	5	1460 103
6	0.10 8949 804	6	2546 987
7	0.11 0007 165	7	3634 498
8	1065 107	8	4722 635
0.10 9	2123 631	0.15 9	5811 400
0.11 0	0.11 3182 737	0.16 0	0.16 6900 794
1	4242 427	1	7990 818
2	5302 702	2	0.16 9081 472
3	6363 562	3	0.17 0172 759
4	7425 009	4	1264 678
5	8487 042	5	2357 230
6	0.11 9549 663	6	3450 418
7	0.12 0612 874	7	4544 241
8	1676 674	8	5638 701
0.11 9	2741 064	0.16 9	6733 799
0.12 0	0.12 3806 046	0.17 0	0.17 7829 536
1	4871 621	1	0.17 8925 912
2	5937 788	2	0.18 0022 929
3	7004 550	3	1120 588
4	8071 907	4	2218 890
5	0.12 9139 860	5	3317 836
6	0.13 0208 410	6	4417 426
7	1277 557	7	5517 663
8	2347 303	8	6618 546
0.12 9	3417 649	0.17 9	7720 078
0.13 0	0.13 4488 595	0.18 0	0.18 8822 258
1	5560 143	1	0.18 9925 089
2	6632 293	2	0.19 1028 570
3	7705 046	3	2132 704
4	8778 403	4	3237 490
5	0.13 9852 365	5	4342 931
6	0.14 0926 933	6	5449 028
7	2002 108	7	6555 780
8	3077 891	8	7663 190
0.13 9	4154 282	0.18 9	8771 259
0.14 0	0.14 5231 283	0.19 0	0.19 9879 987
1	6308 895	1	0.20 0989 375
2	7387 119	2	2099 425
3	8465 954	3	3210 138
4	0.14 9545 404	4	4321 515
5	0.15 0625 467	5	5433 556
6	1706 146	6	6546 264
7	2787 441	7	7659 639
8	3869 354	8	8773 681
0.14 9	4951 884	0.19 9	0.20 9888 393
0.15 0	0.15 6035 034	0.20 0	0.21 1003 775

TABLE II. (cont.)

z		$\zeta(1, 2 z) = -Rl(1-z)$		z		$\zeta(1, 2 z) = -Rl(1-z)$	
0.20	0	0.21	1003 775	0.25	0	0.26	7652 639
	1		2119 829		1		8803 733
	2		3236 555		2	0.26	9955 558
	3		4353 955		3	0.27	1108 116
	4		5472 030		4		2261 409
	5		6590 780		5		3415 437
	6		7710 207		6		4570 202
	7		8830 312		7		5725 705
	8	0.21	9951 097		8		6881 948
0.20	9	0.22	1072 562	0.25	9		8038 931
0.21	0	0.22	2194 708	0.26	0	0.27	9196 656
	1		3317 537		1	0.28	0355 124
	2		4441 049		2		1514 337
	3		5565 246		3		2674 296
	4		6690 130		4		3835 002
	5		7815 700		5		4996 457
	6	0.22	8941 958		6		6158 661
	7	0.23	0068 906		7		7321 617
	8		1196 545		8		8485 325
0.21	9		2324 875	0.26	9	0.28	9649 787
0.22	0	0.23	3453 898	0.27	0	0.29	0815 005
	1		4583 615		1		1980 979
	2		5714 027		2		3147 710
	3		6845 136		3		4315 201
	4		7976 942		4		5483 453
	5	0.23	9109 446		5		6652 466
	6	0.24	0242 651		6		7822 243
	7		1376 557		7	0.29	8992 785
	8		2511 165		8	0.30	0164 092
0.22	9		3646 476	0.27	9		1336 168
0.23	0	0.24	4782 492	0.28	0	0.30	2509 012
	1		5919 213		1		3682 626
	2		7056 642		2		4857 011
	3		8194 779		3		6032 170
	4	0.24	9333 625		4		7208 103
	5	0.25	0473 182		5		8384 812
	6		1613 450		6	0.30	9562 299
	7		2754 432		7	0.31	0740 563
	8		3896 128		8		1919 608
0.23	9		5038 539	0.28	9		3099 434
0.24	0	0.25	6181 667	0.29	0	0.31	4280 044
	1		7325 514		1		5461 437
	2		8470 079		2		6643 616
	3	0.25	9615 364		3		7826 583
	4	0.26	0761 371		4	0.31	9010 338
	5		1908 102		5	0.32	0194 883
	6		3055 556		6		1380 219
	7		4203 735		7		2566 349
	8		5352 642		8		3753 273
0.24	9		6502 276	0.29	9		4940 993
0.25	0	0.26	7652 639	0.30	0	0.32	6129 510

TABLE II. (cont.)

z	$\zeta(1, 2 z) = -Rl(1-z)$	z	$\zeta(1, 2 z) = -Rl(1-z)$
0.30 0	0.32 6129 510	0.35 0	0.38 6605 941
1	7318 826	1	7837 189
2	8508 943	2	0.38 9069 318
3	0.32 9699 861	3	0.39 0302 330
4	0.33 0891 583	4	1536 225
5	2084 109	5	2771 007
6	3277 442	6	4006 676
7	4471 582	7	5243 235
8	5666 532	8	6480 686
0.30 9	6862 293	0.35 9	7719 029
0.31 0	0.33 8058 866	0.36 0	0.39 8958 267
1	0.33 9256 252	1	0.40 0198 402
2	0.34 0454 454	2	1439 436
3	1653 474	3	2681 370
4	2853 311	4	3924 206
5	4053 969	5	5167 946
6	5255 448	6	6412 592
7	6457 750	7	7658 145
8	7660 877	8	0.40 8904 608
0.31 9	0.34 8864 830	0.36 9	0.41 0151 982
0.32 0	0.35 0069 611	0.37 0	0.41 1400 269
1	1275 221	1	2649 471
2	2481 662	2	3899 591
3	3688 936	3	5150 629
4	4897 043	4	6402 587
5	6105 987	5	7655 468
6	7315 768	6	0.41 8909 274
7	8526 387	7	0.42 0164 006
8	0.35 9737 847	8	1419 666
0.32 9	0.36 0950 150	0.37 9	2676 256
0.33 0	0.36 2163 296	0.38 0	0.42 3933 778
1	3377 287	1	5192 235
2	4592 125	2	6451 627
3	5807 812	3	7711 957
4	7024 349	4	0.42 8973 227
5	8241 738	5	0.43 0235 438
6	0.36 9459 981	6	1498 594
7	0.37 0679 079	7	2762 695
8	1899 034	8	4027 743
0.33 9	3119 847	0.38 9	5293 742
0.34 0	0.37 4341 521	0.39 0	0.43 6560 692
1	5564 056	1	7828 595
2	6787 455	2	0.43 9097 454
3	8011 720	3	0.44 0367 271
4	0.37 9236 851	4	1638 048
5	0.38 0462 851	5	2909 786
6	1689 721	6	4182 488
7	2917 464	7	5456 156
8	4146 080	8	6730 792
0.34 9	5375 572	0.39 9	8006 397
0.35 0	0.38 6605 941	0.40 0	0.44 9282 974

TABLE II. (cont.)

z		$\zeta(1, 2 z) = -\text{Rl}(1-z)$			z		$\zeta(1, 2 z) = -\text{Rl}(1-z)$		
0.40	0	0.44	9282	974	0.45	0	0.51	4398	989
	1	0.45	0560	526		1		5728	060
	2		1839	053		2		7058	222
	3		3118	559		3		8389	477
	4		4399	045		4	0.51	9721	828
	5		5680	513		5	0.52	1055	277
	6		6962	966		6		2389	826
	7		8246	405		7		3725	479
	8	0.45	9530	833		8		5062	238
0.40	9	0.46	0816	252	0.45	9		6400	106
0.41	0	0.46	2102	664	0.46	0	0.52	7739	085
	1		3390	071		1	0.52	9079	177
	2		4678	476		2	0.53	0420	386
	3		5967	880		3		1762	713
	4		7258	285		4		3106	163
	5		8549	694		5		4450	736
	6	0.46	9842	109		6		5796	437
	7	0.47	1135	533		7		7143	267
	8		2429	967		8		8491	230
0.41	9		3725	413	0.46	9	0.53	9840	327
0.42	0	0.47	5021	875	0.47	0	0.54	1190	562
	1		6319	353		1		2541	937
	2		7617	851		2		3894	456
	3	0.47	8917	370		3		5248	121
	4	0.48	0217	913		4		6602	934
	5		1519	482		5		7958	898
	6		2822	079		6	0.54	9316	017
	7		4125	707		7	0.55	0674	292
	8		5430	368		8		2033	728
0.42	9		6736	065	0.47	9		3394	325
0.43	0	0.48	8042	799	0.48	0	0.55	4756	089
	1	0.48	9350	572		1		6119	020
	2	0.49	0659	388		2		7483	122
	3		1969	248		3	0.55	8848	399
	4		3280	155		4	0.56	0214	852
	5		4592	112		5		1582	484
	6		5905	120		6		2951	299
	7		7219	181		7		4321	300
	8		8534	300		8		5692	489
0.43	9	0.49	9850	476	0.48	9		7064	869
0.44	0	0.50	1167	714	0.49	0	0.56	8438	444
	1		2486	016		1	0.56	9813	215
	2		3805	383		2	0.57	1189	187
	3		5125	819		3		2566	362
	4		6447	326		4		3944	743
	5		7769	905		5		5324	333
	6	0.50	9093	561		6		6705	135
	7	0.51	0418	294		7		8087	153
	8		1744	109		8	0.57	9470	388
0.44	9		3071	006	0.49	9	0.58	0854	845
0.45	0	0.51	4398	989	0.50	0	0.58	2240	526

BEHAVIOUR OF THE FUNCTION FOR COMPLEX VALUES OF THE ARGUMENT.

8. The values of $f(z)$, when z is complex, can be easily split into their real and imaginary parts. Writing $z=r \exp i\theta$, $\zeta=\rho \exp i\theta$, where $0<\theta<2\pi$, in (2.1), we find *

$$\begin{aligned} f(re^{i\theta}) &= \int_0^r \frac{-\log(1-\rho e^{i\theta}) d\rho}{\rho} \\ &= \frac{1}{2} \int_0^r \frac{-\log(1-2\rho \cos \theta + \rho^2)}{\rho} d\rho + i \int_0^r \arctan \frac{\rho \sin \theta}{1-\rho \cos \theta} \frac{d\rho}{\rho}, \end{aligned} \quad (8.1)$$

where that value of \arctan is taken which vanishes with ρ . A similar expression is obtained by deforming the path of integration into the straight line from 0 to $-r$, and the circle $|z|=r$ from $-r$ to $r \exp i\theta$. Writing $\zeta=r \exp i\phi$, we find

$$\begin{aligned} f(re^{i\theta}) &= f(-r) + \int_{\pi}^{\theta} -\log(1-re^{i\phi}) i d\phi \\ &= f(-r) - \int_{\pi}^{\theta} \arctan \frac{r \sin \phi}{1-r \cos \phi} d\phi + \frac{1}{2} i \int_{\pi}^{\theta} -\log(1-2r \cos \phi + r^2) d\phi, \end{aligned} \quad (8.2)$$

where that value of \arctan is taken which vanishes at the lower limit of integration. Comparison of (8.1) and (8.2) yields two curious integral transformations: also by making θ tend to 0 and 2π in (8.2) we obtain the results

$$\int_0^{\pi} \arctan \frac{r \sin \phi}{1-r \cos \phi} d\phi = f(r) - f(-r), \quad (8.3)$$

where that value of \arctan is taken which vanishes at the upper limit, and

$$\left. \begin{aligned} \int_0^{\pi} \log(1-2r \cos \phi + r^2) d\phi &= 0, & r \leq 1 \\ &= 2\pi \log r, & r \geq 1 \end{aligned} \right\} \quad (8.4)$$

* The imaginary part can be expressed in terms of an integral involving a single parameter (Kummer 1840). Writing

$$\frac{1}{2}u = \arctan \frac{\rho \sin \theta}{1-\rho \cos \theta}, \quad \frac{1}{2}u_1 = \arctan \frac{r \sin \theta}{1-r \cos \theta}, \quad \rho = \frac{\sin \frac{1}{2}u}{\sin(\frac{1}{2}u + \theta)},$$

we have

$$\begin{aligned} \int_0^r \arctan \frac{\rho \sin \theta}{1-\rho \cos \theta} \frac{d\rho}{\rho} &= \int_{\rho=0}^{\rho=r} \frac{1}{2}u d(\log \rho) \\ &= \left[\frac{1}{2}u \log \rho \right]_{\rho=0}^{\rho=r} - \frac{1}{2} \int_{\rho=0}^{\rho=r} \log \rho du \\ &= \frac{1}{2}u_1 \log r + \frac{1}{2} \{Cl(u_1) - Cl(0) - Cl(u_1 + 2\theta) + Cl(2\theta)\}. \end{aligned}$$

Here

$$Cl(x) = - \int_0^x \log(2 \sin \frac{1}{2}x) dx$$

is the integral defined and tabulated by Clausen (1828), for $x=0^\circ(1^\circ)180^\circ:16D$.

9. More interesting definite integrals are obtained by using the various formulæ (3.3), (4.1), etc., to determine $f(z)$ as far as possible, for special complex values of z , and comparing with (8.1) and (8.2). Putting $z=\frac{1}{2}+iy$ in (3.3), we find

$$\Re f(\tfrac{1}{2}+iy) = \frac{\pi^2}{12} - \tfrac{1}{8} \log^2 (\tfrac{1}{4}+y^2) - \tfrac{1}{2} \arctan^2 2y, \quad . \quad . \quad (9.1)$$

where $|\arctan 2y| < \frac{1}{2}\pi$ and the prefix \Re indicates the real part. This gives us, for any acute angle θ ,

$$\int_0^{\frac{1}{2}\sec\theta} \frac{-\log(1-2\rho\cos\theta+\rho^2)}{\rho} d\rho = \frac{\pi^2}{6} - \theta^2 - \log^2(2\cos\theta), \quad . \quad (9.2)$$

$$\int_0^\pi \arctan\left(\frac{\frac{1}{2}\sec\theta\sin\phi}{1-\frac{1}{2}\sec\theta\cos\phi}\right) d\phi = \frac{\pi^2}{6} - \theta^2 - \log^2(2\cos\theta) - 2f(-\tfrac{1}{2}\sec\theta). \quad . \quad . \quad . \quad (9.3)$$

Again, putting $z=\exp i\theta$, where $|\theta|<\pi$, in (4.1),

$$\Re f(e^{i\theta}) = \tfrac{1}{8}\pi^2 - \tfrac{1}{2}\pi|\theta| + \tfrac{1}{4}\theta^2. \quad . \quad . \quad . \quad (9.4)$$

Hence, for any θ numerically less than π ,

$$\int_0^1 \frac{-\log(1-2\rho\cos\theta+\rho^2)}{\rho} d\rho = \tfrac{1}{8}\pi^2 - \pi|\theta| + \tfrac{1}{2}\theta^2, \quad . \quad . \quad (9.5)$$

the integral derived from (8.2) and (9.4) being trivial.

Putting $z=1+\exp i\theta$, where $|\theta|\leq\pi$, in (5.2), we find after some reduction

$$\Re f(1+e^{i\theta}) = \tfrac{1}{4}\pi^2 + \tfrac{1}{4}\theta^2 - \tfrac{1}{2}\pi|\theta|. \quad . \quad . \quad . \quad (9.6)$$

Applying (8.1) and (8.2), for any acute angle θ ,

$$\int_0^{2\cos\theta} \frac{-\log(1-2\rho\cos\theta+\rho^2)}{\rho} d\rho = \tfrac{1}{2}\pi^2 + 2\theta^2 - 2\pi|\theta|, \quad . \quad . \quad (9.7)$$

$$\int_0^\pi \arctan \frac{2\cos\theta\sin\phi}{1-2\cos\theta\cos\phi} d\phi = \tfrac{1}{4}\pi^2 + \theta^2 - \pi|\theta| - f(-2\cos\theta). \quad . \quad (9.8)$$

Finally, putting $z=iy$ in (6.1) and $z=r\exp i\pi/3$, $r\exp 2i\pi/3$ in (6.2) with $n=3$, we obtain the relations

$$\left. \begin{aligned} \Re f(iy) &= \tfrac{1}{4}f(-y^2) \\ \Re f(re^{i\pi/3}) &= \tfrac{1}{6}f(-r^3) - \tfrac{1}{2}f(-r) \\ \Re f(re^{2i\pi/3}) &= \tfrac{1}{6}f(r^3) - \tfrac{1}{2}f(r) \end{aligned} \right\}. \quad . \quad . \quad . \quad (9.9)$$

Applying (8.1), these give us

$$\left. \begin{aligned} \int_0^y \frac{-\log(1+\rho^2)}{\rho} d\rho &= \tfrac{1}{2}f(-y^2) \\ \int_0^r \frac{-\log(1-\rho+\rho^2)}{\rho} d\rho &= \tfrac{1}{3}f(-r^3) - f(-r) \\ \int_0^r \frac{-\log(1+\rho+\rho^2)}{\rho} d\rho &= \tfrac{1}{3}f(r^3) - f(r) \end{aligned} \right\},$$

all trivial relations obtainable more easily by other methods. On the other hand (8.2) gives

$$\left. \begin{aligned} \int_{\pi/2}^{\pi} \arctan\left(\frac{y \sin \phi}{1-y \cos \phi}\right) d\phi &= \frac{1}{2}f(-y^2) - f(-y) \\ \int_{\pi/3}^{\pi} \arctan\left(\frac{r \sin \phi}{1-r \cos \phi}\right) d\phi &= \frac{1}{6}f(-r^3) - \frac{3}{2}f(-r) \\ \int_{2\pi/3}^{\pi} \arctan\left(\frac{r \sin \phi}{1-r \cos \phi}\right) d\phi &= \frac{1}{6}f(r^3) - \frac{1}{2}f(r) - f(-r) \end{aligned} \right\}. \quad (9.10)$$

A GENERALIZATION OF THE FUNCTION $f(z)$, AND A SUGGESTED NOTATION.

10. Making the substitution

$$\zeta = ze^{-u}, \quad (10.1)$$

in (2.1), we have

$$f(z) = \int_0^{\infty} -\log(1-ze^{-u}) du.$$

Integrating by parts, this becomes

$$f(z) = z \int_0^{\infty} \frac{ue^{-u} du}{1-ze^{-u}}, \quad (10.2)$$

the integrated part vanishing at both limits. This gives us an integral representation in which the integrand is a single-valued function, and definition by this integral (the Cauchy principal value being understood when z is real and positive) agrees with that of §2.

The resemblance of (10.2) to the well-known integral representation of the Riemann ζ -function suggests a generalization of the function $f(z)$, and a notation for it and its generalizations. For values of z of modulus less than unity we may define the function $\zeta(a, s | z)$, the generalization considered, by the ascending series

$$\zeta(a, s | z) = \sum_{n=0}^{\infty} \frac{z^{n+1}}{(a+n)^s}. \quad (10.3)$$

This sum can be replaced by the infinite integral *

$$\zeta(a, s | z) = \frac{z}{\Gamma(s)} \int_0^{\infty} \frac{u^{s-1} e^{-au} du}{1-ze^{-u}}, \quad (10.4)$$

where $\arg u = 0$, and this defines the function for all values of z , if the real parts of a and $s-1$ are both greater than zero, and the principal value is understood if z is real and $z \geq 1$. This in its turn may be replaced by the complex integral

$$\zeta(a, s | z) = -z \frac{\Gamma(1-s)}{2\pi i} \int_{\infty}^{(0+)} \frac{(-u)^{s-1} e^{-au}}{1-ze^{-u}} du, \quad . . . (10.5)$$

* The Fermi-Dirac functions (McDougall and Stoner 1938), defined by $F_k(\eta) = \int_0^{\infty} \frac{x^k dx}{e^x - 1}$, are special cases: we find $F_k(\eta) = -\Gamma(k+1) \zeta(1, k+1 | -e^{-\eta})$.

which extends the definition to all values of s . The path of integration here is from $+\infty$, along the real axis, around the origin in the positive sense, and back along the real axis to $+\infty$, the contour being deformed if necessary so that none of the poles of the integrand (the points $u = \log z + 2n\pi i$) lies within it. The proofs of these statements so exactly parallel those for the Riemann ζ -function that there is no need to quote them. The only modification, due to the presence of z in the denominator, is that for real values of z greater than unity the contour must describe, once in each direction, a small semicircle above or below the pole $u = \log z$: the arithmetic mean of the values so obtained agrees with the definition of §2, for the particular case there considered.

Transformed by the substitution (10.1), (10.4) becomes

$$\zeta(a, s | z) = \frac{1}{\Gamma(s)} \int_0^z \frac{(\zeta/z)^{a-1} \{-\log(\zeta/z)\}^{s-1} d\zeta}{1-\zeta}, \quad \dots \quad (10.6)$$

a general logarithmic integral.

11. This proposed generalization arises naturally, for $n=1$ and integral values of s , when we consider integrals involving the function $f(z)$, or $\zeta(1, 2 | z)$, under the sign of integration. Thus, integrating by parts,

$$\int_0^x f(y) dy = \left[yf(y) \right]_0^x + \int_0^x \log(1-y) dy = xf(x) - (1-x) \log(1-x) - x. \quad (11.1)$$

Also, if n is not equal to -1 ,

$$\begin{aligned} \int_0^x y^n f(y) dy &= \left[y^n \{ yf(y) - (1-y) \log(1-y) - y \} \right]_0^x \\ &\quad - \int_0^x ny^{n-1} \{ yf(y) - (1-y) \log(1-y) - y \} dy, \end{aligned}$$

whence

$$\begin{aligned} \int_0^x y^n f(y) dy &= \left[\frac{y^n}{n+1} \{ yf(y) - (1-y) \log(1-y) - y \} \right]_0^x \\ &\quad + \frac{n}{n+1} \int_0^x y^{n-1} \{ (1-y) \log(1-y) + y \} dy. \quad \dots \quad (11.2) \end{aligned}$$

Two further integrations by parts leave us with an integral whose integrand is algebraic. For the exceptional value $n=-1$,

$$\begin{aligned} \int_0^x \frac{f(y)}{y} dy &= \left[f(y) \log \frac{y}{x} \right]_0^x + \int_0^x \frac{\log \frac{y}{x} \log(1-y)}{y} dy \\ &= \left[\frac{1}{2} \log^2 \frac{y}{x} \log(1-y) \right]_0^x + \frac{1}{2} \int_0^x \frac{\log^2 \frac{y}{x}}{1-y} dy \\ &= \zeta(1, 3 | z), \quad \dots \quad (11.3) \end{aligned}$$

illustrating how $\zeta(1, 3 | z)$ arises,

A simple generalization of this result,

$$\int_0^x \frac{\zeta(1, s-1|y)}{y} dy = \zeta(1, s|x), \quad \dots \quad (11.4)$$

valid if $s > 1$, is easily proved by differentiation of (10.6). More generally, we have the reduction formulæ

$$\int_0^x y^n \zeta(a, s|y) dy = \frac{1}{n-a+2} \left\{ x^{n+1} \zeta(a, s|x) - \int_0^x y^n \zeta(a, s-1|y) dy \right\}, \quad (11.5)$$

and

$$\int_0^x \frac{\zeta(a, s|y)}{y} dy = -\frac{1}{a-1} \left\{ \zeta(a, s|x) - \int_0^x \frac{\zeta(a, s-1|y)}{y} dy \right\}, \quad (11.6)$$

valid if the real part of s is greater than unity. Both formulæ are easily established by differentiation of (10.6).

REFERENCES.

- CLAUSEN, T., 1832, *J.f.d. reine u. angew. Math. (Crelle)*, **8**, 298-300.
 FLETCHER, A., 1944, *Phil. Mag. [7]*, **35**, 16-17.
 HILL, C. J., 1828, *J.f.d. reine u. angew. Math. (Crelle)*, **3**, 101-159; *Specimen exercitii analytici functionem integrelem* $\int \frac{dx}{x} \log(1+2x \cos \alpha + x^2)$ *tum quoad amplitudinem tum quoad modulum comparandi modum exhibenſis* (Londini Gothorum 1830).
 KUMMER, E. E., 1840, *J.f.d. reine u. angew. Math. (Crelle)*, **21**, 74-90, 193-225, 328-371.
 MCDUGALL, J. and STONER, E. C., 1938, *Phil. Trans. No. 773*, **237**, 67-104.
 NEWMAN, F. W., 1892, *The Higher Trigonometry. Superrationals of Second Order* (Cambridge).
 POWELL, E. O., 1943, *Phil. Mag. [7]*, **34**, 600-607.
 SPENCE, W., 1809, *An essay on the theory of the various orders of logarithmic transcendents* (London and Edinburgh).

MARCONI MEASURING INSTRUMENT EXHIBITION.

MARCONI INSTRUMENTS LIMITED are showing a wide range of Industrial Measuring Instruments at their London Showroom, 109 Eaton Square, London, S.W.1. The Exhibition will be open for two weeks, March 14th to 25th (Mon. to Fri.) and all interested in electronics and process control are cordially invited to apply for admission tickets.

Among the exhibits will be the latest pH Meters and Moisture Meters, a range of Industrial Recorders and a number of general-purpose instruments for widely differing applications. There will be many working models.

Tickets may be obtained either from Marconi Instruments Ltd., St. Albans, Herts., or the London Showroom.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XXXIII. *Ionization and Charge Exchange by Fast Ions of Hydrogen and Helium.*

By J. P. KEENE,
Cavendish Laboratory, Cambridge*.

[Received December 13, 1948.]

ABSTRACT.

An account is given of experiments involving collisions of H^+ , H_2^+ , and He^+ ions with hydrogen and helium in the energy range 3–35 KeV. The construction of apparatus for producing suitable beams of ions is described and details presented of an accurate method for measuring charge exchange and ionization cross-sections. Curves showing values of these cross-sections are given and they are compared with the results of other workers where these are available. A method for measuring e/m values of secondary ions is also described.

I. INTRODUCTION.

THE processes which occur when positive ions collide with gas atoms or molecules have been the subject of numerous investigations. Elastic scattering, charge exchange, ionization, excitation, dissociation, and onset potentials have been studied in many cases, and incident energies have ranged from a few up to millions of electron-volts. Although a large number of papers have been published on the subject a great deal of work remains to be done. Discrepancies exist between the results of various workers and it is clear that the effects of secondary processes have not always been considered, thus cross-section measurements may not have been reliable.

The present paper is concerned with the measurement of charge exchange and ionization cross-sections for H^+ , H_2^+ , and He^+ ions in hydrogen and helium in the energy range 3–35 KeV. The main contribution to the total collision cross-section is charge exchange as elastic scattering is mostly through very small angles (Smith 1934 a, b).

Charge exchange may be represented in the simplest case by :



The incident ion I^+ continues as a neutral particle with only small change of direction or velocity. It leaves behind it a slow ion G^+ of the gas G through which it is passing.

* Communicated by Mr. E. S. Shire.

Ionization usually occurs with a smaller cross-section and is represented by :



A slow electron e^- is produced in addition to a slow ion and the original fast ion continues as a charged particle.

Measurements of charge exchange cross-sections by various other workers do not agree in many cases, while ionization cross-sections have seldom been measured.

The method used in the experiments described here was similar in principle to that used by Goldmann (1931) for 0.5–4.0 KeV. H^+ ions in hydrogen, helium, and argon, and by Sherwin (1940) for 6–24 KeV. metallic ions in hydrogen and helium. A parallel homogeneous beam of ions passed through the gas, the pressure being such that only a small fraction of the beam was lost by collisions. Ions or electrons formed along its path were extracted by a transverse electrostatic field and the saturation currents measured. The corresponding cross-sections were calculated from the values of these currents, together with the values of pressure, beam current, and path length. In the present experiments a thorough investigation of all secondary electron effects was also made.

The collision processes in equations (1) and (2) are the simplest type and there is evidence that they are usually predominant. Cross-section measurements, however, cannot be regarded as satisfactory without some accompanying information as to the nature of the collision products. In the experiments described here a magnetic analysis of the slow ions was made in order to supply this information.

II. APPARATUS.

As most of the techniques associated with the production of beams of positive ions are fairly well known only a general description of the apparatus will be given.

Fig. 1 is a schematic diagram of the vacuum system. Ions were produced in the source s which consisted of an axial low voltage arc running in a magnetic field between a hot tungsten filament and a water-cooled anode. The filament was U-shaped with its plane at right angles to the axis of the system; it was placed immediately above the exit hole h_1 , and some of the ions from the arc passed through the middle of the U and through h_1 , emerging with very little energy spread. The resulting divergent beam of ions was focused into a parallel beam, 3–4 mm. in diameter, by the two electrostatic lenses l_1 and l_2 . The ions required were selected by the magnetic field in m and passed through h_2 and h_3 , into the measuring tube g where collisions with gas atoms were investigated.

Most of the vacuum system was constructed of metal with double neoprene gasket seals between the various parts. The tubes t_1 and t_2 each led to a 20 litre/sec. mercury diffusion pump via a liquid air trap. Pressures in s , m , and b were measured with an ionization gauge previously calibrated against a McLeod gauge. A separate ionization gauge and liquid air trap were connected to g . Pressures in m and b were found

to be of the order of 3×10^{-6} mm./Hg when no gas was being supplied to *s* and *g*. When the gas supplies were turned on these pressures rose to about 10^{-5} mm./Hg. The pumping system was arranged so that the amount of gas from *s* which could reach *g* was extremely small.

The electrode systems used in the experiments were supported inside the glass tube *g* by means of a removable ground glass cone. The details of *g* are not shown in the diagram, and it is only necessary to mention the side tubes, which led to an ionization gauge, a liquid air trap, and a gas inlet. The gas used was cleaned, by passing it over charcoal cooled in liquid air, immediately before it entered the vacuum system.

Fig. 1.

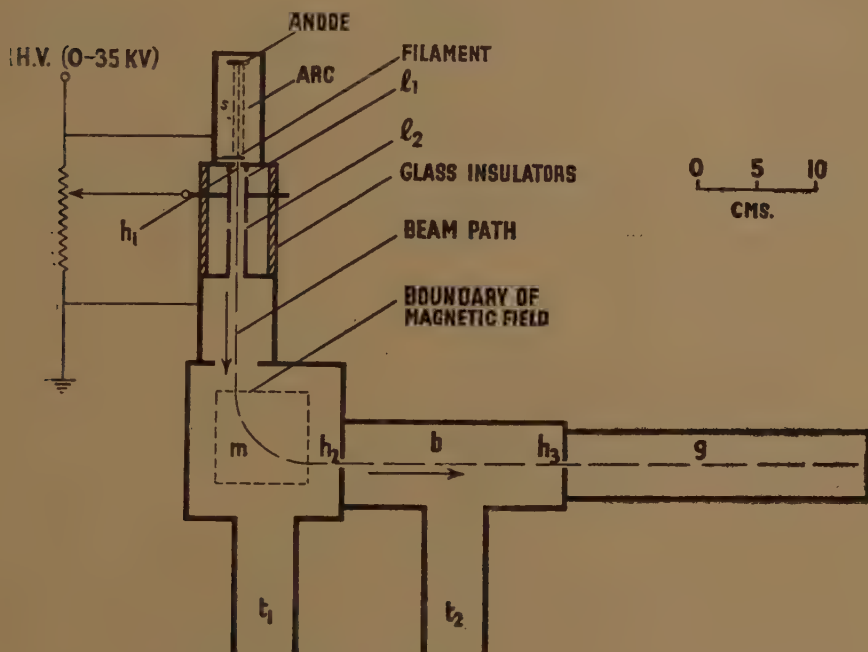


Diagram of vacuum system.

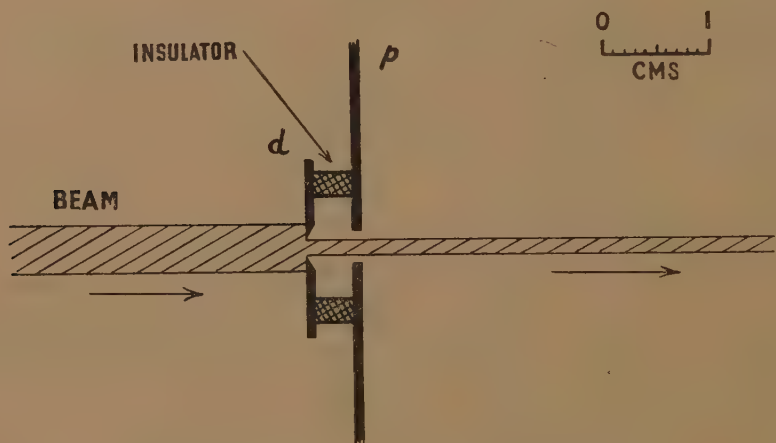
The ion source produced total beam currents of about $30 \mu\text{A}$, through the exit hole. Resolved currents of H_2^+ and He^+ up to about $1.0 \mu\text{A}$, and of H^+ up to $0.2 \mu\text{A}$, could be obtained in *g* when the defining holes h_2 and h_3 were in place (h_3 was 1.5 mm. diameter). The beam current was measured with a galvanometer, and other currents with an electrometer tetrode in the usual balanced D.C. amplifying circuit (Penick 1935). Input resistances up to $10^{10} \omega$ were used so that precautions had to be taken against leakage, particularly as potentials as large as 60 v. between electrodes were involved. The lead carrying the small current was screened wherever possible by a tube maintained at the same potential, and all insulators were coated with ceresin wax. No difficulty was experienced in measuring currents down to 10^{-13} amp. with an accuracy of ± 5 per cent.

III. EXPERIMENTS.

In preliminary experiments it was found that secondary electrons from the edge of h_3 and charge exchange ions ejected from the Faraday cage, were being included in the measured currents to the collecting plate; there were also indications of the presence of other secondary electrons.

The original hole h_3 was a plain hole in $1/32$ in. sheet metal and secondary electrons produced at its edge were entering g ; in order to eliminate this effect the arrangement shown in fig. 2 was used. The original hole in p was enlarged to 3 mm. and a disc d mounted behind it on an insulating ring. The beam was defined by the 1.5 mm. hole in d (which had a sharp edge) and cleared the edge of the hole in p . A positive potential of 30 v. was applied to d , p being earthed, so that secondary electrons formed at the edge of the hole in d could not enter g .

Fig. 2.



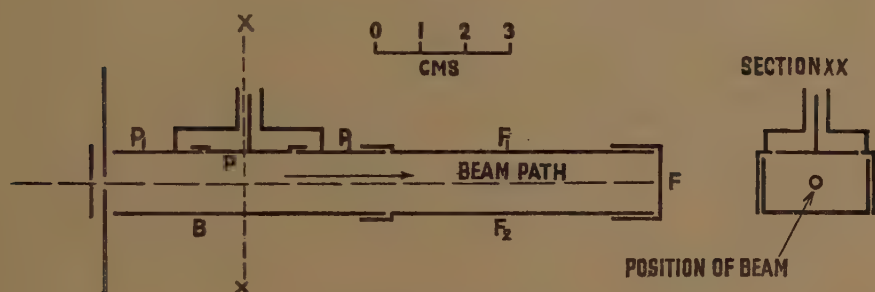
The hole through which the beam entered g .

The first Faraday cage to be used consisted of a long cylinder closed at one end and maintained at $+30$ v. to prevent the escape of secondary electrons. Under these conditions charge exchange ions, formed with energies of a few electron volts inside the cage, were accelerated outwards and considerably increased the current to the collecting plate. To overcome this the Faraday cage was arranged so that a transverse electric field inside it trapped slow ions as well as electrons (see fig. 3).

The experiments described above suggested the design of two electrode systems for the final measurements. Electrode system 1 is shown schematically in fig. 3. The various parts were made of Ferry wire and sheet, construction being by spot welding. Each electrode was supported on a 1.5 mm. Ferry wire leading to a steatite ring carried by the cone in g . The Faraday cage (made up by F , F_1 and F_2) collected the beam, and 30 v. was applied between F_1 and F_2 . The field between F_1 and

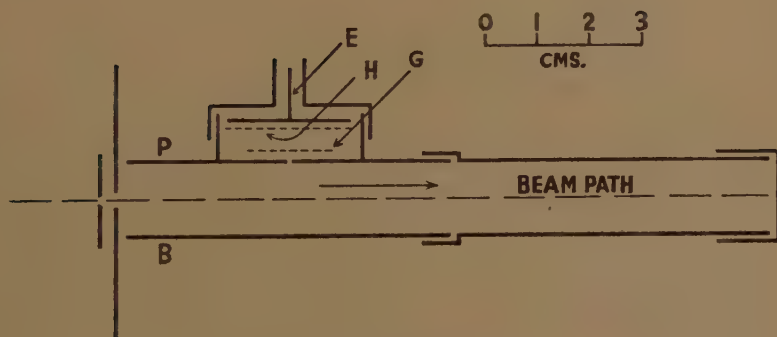
F_2 was always arranged to be in the same direction as that between P and B. F was connected to F_1 or F_2 , whichever was positive, and the total current to the three electrodes was measured. The two guard electrodes P_1 were at the same potential as the collecting electrode P and the lead to P was screened as shown in the diagram. B and F_2 were usually both at zero potential and a suitable voltage was applied to P_1PP_1 . In this way a uniform transverse field was obtained and ions or electrons formed in a length of beam path equal to the exposed length of P were collected on P (this length was 2 cm.).

Fig. 3.



Electrode system 1.

Fig. 4.



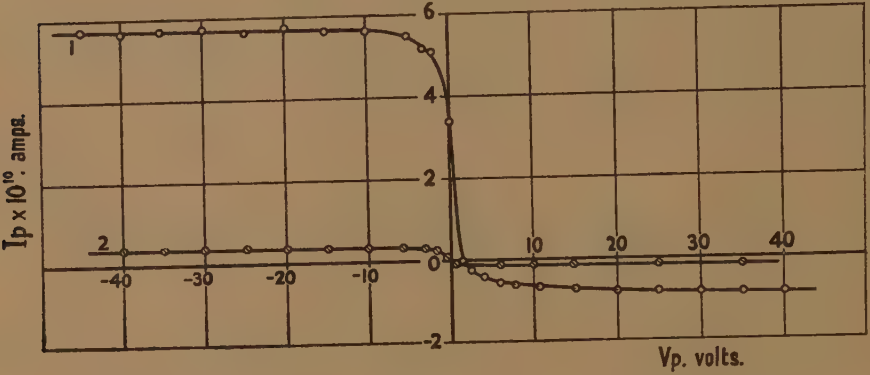
Electrode system 2.

Electrode system 2 is shown in fig. 4. It was designed so that the origin of the currents to P in electrode system 1 could be investigated by measurement of the energy with which the charged particles arrived at P. The plates P_1 and P were replaced by a single plate P in which was a 0.9 mm. slit; behind this slit were two grids G and H and a collecting plate E. The grids consisted of very fine parallel wires about 0.5 mm. apart, the direction of the wires being parallel to the direction of the main beam (for convenience the wires are shown at right angles to this direction in fig. 4). Each grid had a transmission of about 90 per cent.

(i) Collection of slow ions and electrons.

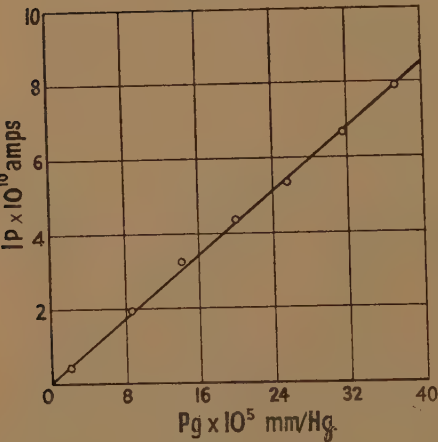
Using electrode system 1 with 15 KeV. H_2^+ ions in hydrogen the current to P (I_P) was measured as a function of the potential on P with respect

Fig. 5.



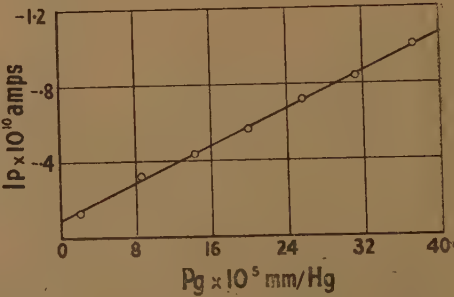
I_P as a function of V_P for 15 KeV. H_2^+ ions. ($I_F 390 \times 10^{-10}$ amp.).
Curve 1. Hydrogen pressure 28×10^{-5} mm./Hg.
Curve 2. Hydrogen supply turned off.
Pressure 1.9×10^{-5} mm./Hg.

Fig. 6.



I_P as a function of P_g for 15 KeV. H_2^+ ions in hydrogen.
 $I_F 390 \times 10^{-10}$ amp. $V_P - .15$ v.

Fig. 7.



I_P as a function of P_g for 15 KeV. H_2^+ ions in hydrogen.
 $I_F 390 \times 10^{-10}$ amp. $V_P + 25$ v.

current. With a hydrogen pressure of 28×10^{-5} mm./Hg. I_P was of the order of 10^{-10} amp. for a beam current (I_F) of 390×10^{-10} amp. The

results are shown in fig. 5 together with the corresponding curve for the case when no hydrogen was being supplied to g ; the points on the graph are reduced to a standard value of I_F .

For fixed values of V_P and of gas pressure (P_g) it was found that I_P was proportional to I_F , and for fixed values of V_P and I_F it was proportional to P_g . The latter case is shown in figs. 6 and 7 where I_P is plotted as a function of P_g for $V_P - 15$ v. and $+25$ v. At pressures much greater than 40×10^{-5} mm./Hg. the linear relation ceased to hold. The lines do not quite go through the origin and this may be due to the fact that residual gas was present (pressure 1.9×10^{-5} mm./Hg.); the corresponding lines for this residual gas may be expected to have different slopes.

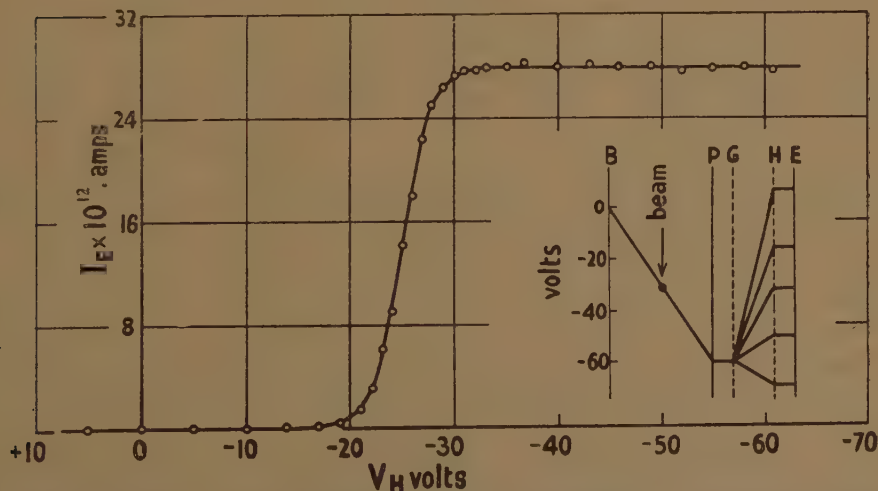
The curves shown in figs. 5, 6 and 7 are typical of all the ions and gases investigated. The positive and negative currents always saturated easily, although the rate of saturation varied with the particular ion and gas.

Before using results of this kind for calculating cross-sections it was necessary to investigate the origin of the currents to P, and the secondary electron emission from various electrodes. These experiments are described next.

(ii) Origin of the measured currents.

Figs. 8, 9 and 10 are examples of the results obtained with electrode system 2. B was always kept at zero potential; V_P and V_H are the

Fig. 8.



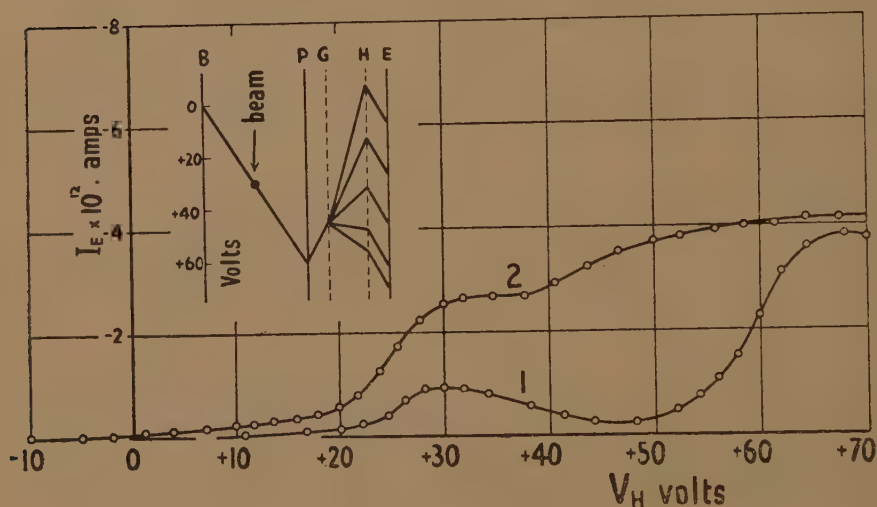
I_E as a function of V_H for 15 KeV. H_2^+ ions in hydrogen. $I_F 390 \times 10^{-10}$ amp. $P_g 34 \times 10^{-5}$ mm./Hg. $V_P - 60$ v. V_G and V_E zero.

potentials of P and H with respect to B; V_G is the potential of G with respect to P, and V_E is the potential of E with respect to H. The diagrams attached to figs. 8 and 9 show the distribution of potential, the vertical

lines representing the electrodes; in each case the force on the particles being considered was such as to make them "run down-hill" on the diagram. V_P was set at ± 60 v. and I_E measured as a function of V_H for fixed gas pressure and potentials on G and E.

Fig. 8 shows the curve for $V_P -60$ v., V_G and V_E zero, P_g 34×10^{-5} mm./Hg., and a beam of 15 KeV. H_2^+ ions in hydrogen. It is evident that the positive ions collected on P were formed in a region whose potential lay between -20 v. and -30 v. Clearly the ions were formed in the beam and the curve shows that the beam was not accurately aligned since the current rose to half value at -25 v. instead of the expected -30 v. As the beam diameter was 1.5 mm., the distance from P to B

Fig. 9.



I_E as a function of V_H for 15 KeV. H_2^+ ions in hydrogen. I_F 390×10^{-10} amp. P_g 34×10^{-5} mm./Hg. $V_P + 60$ v.

Curve 1. V_G and V_E zero. Curve 2. $V_G - 18$ v. $V_E + 18$ v.

14 mm., and the potential drop 60 v., there existed a drop of 6.4 v. across the beam. I_E in fig. 8 rises from 10 per cent to 90 per cent of its maximum value for a change in V_H of 6.4 v. so that the ions must have been formed with very little kinetic energy.

The corresponding curve for electrons is shown in fig. 9 (curve 1), V_P was $+60$ v. in this case, V_G and V_E being zero, and P_g 34×10^{-5} mm./Hg. It is obvious from the shape of this curve that some secondary effects must have been occurring and in order to elucidate them an investigation was made of the variation of I_E with V_G and V_E for several fixed values of V_H . It was found that secondary electrons were being formed on various electrodes and that the shape of curve 1 could be explained in

terms of this secondary emission as follows. (a) The initial rise between 20 v. and 30 v. is due to collection of electrons formed in the beam, and corresponds to the rise in positive ion current shown in fig. 8. (b) The subsequent fall in current is due to the fact that secondary electron emission (increasing in magnitude as V_H increased) was produced at the surface of E by the primary electrons from the beam. Most of the secondaries, assisted by field penetration through H, were able to escape. (c) As V_H increases through 60 v. the curve again rises rapidly because electrons formed on the surface of P were beginning to be collected at E. The arrival of 30 eV. electrons from the beam produced secondary electrons at the surface of P, but generally the field was such as to prevent their escape. Near the slit, however, some secondaries were formed with suitable initial directions and velocities to enable them to pass through it. As most of them had energies of only a few electron-volts they were not collected on E until V_H was in the region of 60 v. (d) The fall in current beyond 60 v. is caused in the same way as the fall beyond 30 v.

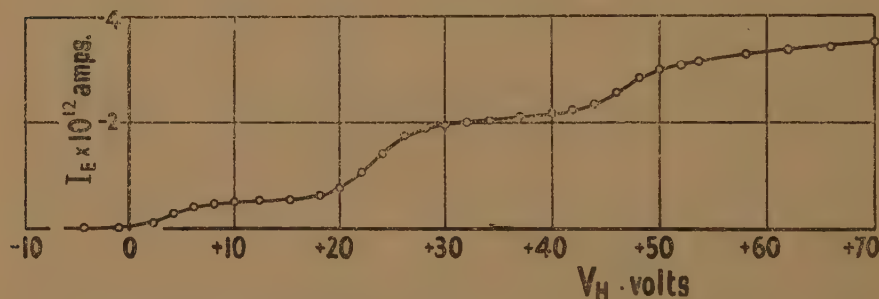
Curve 2 (fig. 9) shows the effect of applying suitable voltages to the grids to suppress the secondary currents. For this curve, V_P was again +60 v., V_E was +18 v. in order to prevent the escape of secondary electrons from E, and V_G -18 v. so that electrons passing through the slit with less than 18 eV. energy were prevented from reaching E. The potential distribution is shown in the attached diagram. The rise beyond 40 v. is due to secondary electrons from P; some of these reached E in spite of the retarding voltage, because secondaries are usually formed with an appreciable energy spread.

The experiments so far described indicate that the currents shown at -35 v. in fig. 8 and +35 v. in fig. 9 (curve 2) were samples of the ions and electrons formed along the beam path. In order to investigate them further the gas pressure was varied and the currents were found to be proportional to pressure. The ratio of the slopes for V_H +35 v. (with V_G -18 v. and V_E +18 v.) and V_H -35 v. agreed within 15 per cent with the ratio of the corresponding slopes for I_P using electrode system 1; better agreement can hardly be expected as I_E was still increasing in the region of 35 v. (see fig. 9, curve 2). In addition the appropriate currents to P and E were found to be in approximately the same ratio as the widths of P and of the slit.

Similar curves to those already described were obtained for H^+ and He^+ ions in hydrogen, and in the energy range 3-35 KeV. no significant change in their general shape occurred. The results all show that the currents to P with electrode system 1 were due to charge exchange and ionization along the beam path. Secondary emission from B by fast ions scattered out of the beam, by slow positive ions, or by photons would have appeared as a sudden increase in current as V_H increased through zero in curve 2, fig. 9. The direct arrival of fast positive ions would also have been detected as a positive current for negative values of V_H on the same curve.

When the experiments were repeated with H^+ , H_2^+ , and He^+ ions in helium, similar curves were obtained for the positive ion currents, but the electron currents behaved differently. A typical example is shown in fig. 10 which is for 15 KeV. H_2^+ ions in helium at a pressure of 30×10^{-5} mm./Hg. V_P was +60 v., V_G -15 v., and V_E +18 v. The increases in I_E at +25 v. and +40 v. are similar to those in fig. 9, but in addition the curve rises through V_H zero and then remains fairly constant until the increase at +25 v. occurs. This new effect is due to electron emission from B which was probably caused by the collection of slow He^+ ions at that electrode. The ionization potential of helium is 24.5 v. as compared with 13.5 v. and 15.4 v. for the hydrogen atom and molecule respectively, so that it is quite likely that He^+ ions would cause secondary electron emission. It is also possible that the secondary emission was due to the metastable state of helium (~ 20 v.), or (less likely), to photons, as collision processes in the beam might have resulted

Fig. 10.



I_E as a function of V_H for 15 KeV. H_2^+ ions in helium. I_F 390×10^{-10} amp.
 P_g 30×10^{-5} mm./Hg. V_P +60 v. V_G -15 v. V_E +18 v.

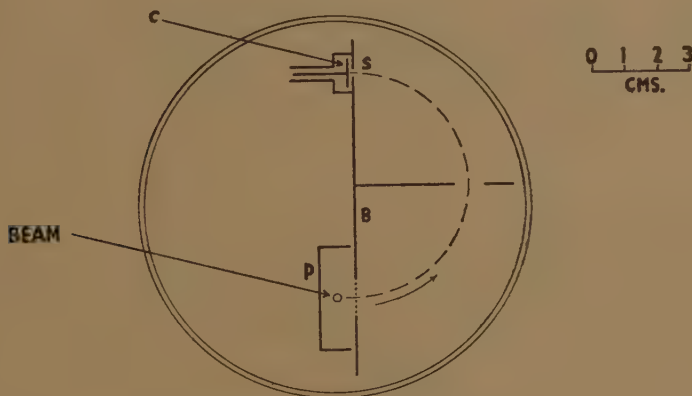
in excitation of helium atoms. The effect cannot have been due to secondary emission by fast scattered ions since the current to E did not go positive for negative values of V_H (fig. 10).

These results for helium show that the electron currents measured to P in electrode system 1 were partly due to electrons from the beam and partly to electrons from the surface of B. The ratio of the currents at +12 v. and +35 v. in curves of the type shown in fig. 10 gave the fraction which had to be subtracted from the saturated electron currents to P (with electrode system 1) when finding the true currents due to ionization in the beam. The procedure with the positive currents was less complicated. These currents were composed of charge exchange ions, ionization ions, and a positive component due to loss of secondary electrons (in this case from P). Since each ionizing collision produced one electron and one ion (see (iii) below) subtraction of the value of the saturated electron current to P (with V_P positive) from the saturated positive current (with V_P negative), gave the positive ion current due to charge exchange.

(iii) *Magnetic analysis.*

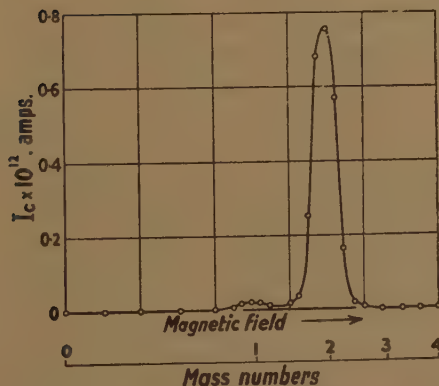
In order to identify the slow ions formed in the beam subsidiary experiments were carried out in which the ions were analysed magnetically. The glass tube was replaced by a brass tube, diameter 13 cm., and the beam passed through this tube parallel to the axis but at a distance of 3.5 cm. from it. A pair of Helmholtz coils provided a uniform magnetic field

Fig. 11.



Magnetic analysis of slow ions.

Fig. 12.



Mass spectrum for 15 KeV. H_2^+ ions in hydrogen. P_0 35×10^{-5} mm./Hg. $V_P + 20$ v.

field (parallel to the axis of the tube) which was used to analyse the slow ions extracted from the beam. The arrangement is shown schematically in fig. 11, the magnetic field being perpendicular to the plane of the diagram. The beam passed between P and B and was collected in a Faraday cage. B was at zero potential and P was made positive so that ions formed in the beam were accelerated towards the grid in B and passed through it into a field free space. Here they were deflected through

180° and collected on C behind the slit S. The arrangement was essentially a mass spectrograph, of the 180° type, in which the beam of electrons in the ion source had been replaced by a beam of fast positive ions. The mass spectra were obtained by varying the magnetic field and measuring the current to C (I_C).

Using H^+ , H_2^+ , and He^+ ions in hydrogen a large peak was found at mass 2, and when the hydrogen was replaced by helium the peak appeared at mass 4. Thus the ions formed in the beam were H_2^+ ions for hydrogen gas and He^+ ions for helium. A very small peak appeared at mass 1 for H^+ and H_2^+ ions in hydrogen, this is shown in fig. 12. The area of the small peak is only a few per cent of that of the main peak, and it indicates that dissociation of the H_2 molecule occurred, the cross-section for the process being very much smaller than the charge exchange cross-section.

IV. RESULTS.

When a beam of positive ions passes through a gas at a pressure such that only a very small fraction of it is lost by collision, the saturated current of slow positive ions formed in a length L of the beam path is given by :

$$I_+ = I_0 N L (\sigma_c + \sigma_i), \quad (3)$$

where I_0 is the beam current,

N the number of gas atoms per cm^3 ,

σ_c the cross-section for charge exchange,

σ_i the cross-section for ionization.

The corresponding current of slow electrons is :

$$I_- = I_0 N L \sigma_i. \quad (4)$$

These results apply only if the assumption is made that the collision processes are predominantly those given by equations (1) and (2). Thus in the case of helium gas, for example, if He^{++} ions were formed in some of the ionizing collisions, equations (3) and (4) would no longer hold. Evidence supporting this assumption has been given in the previous section where it was shown that only H_2^+ and He^+ ions were formed in any quantity.

We therefore have :

$$\sigma_c + \sigma_i = \frac{I_+}{I_0} \cdot \frac{1}{NL}, \quad (5)$$

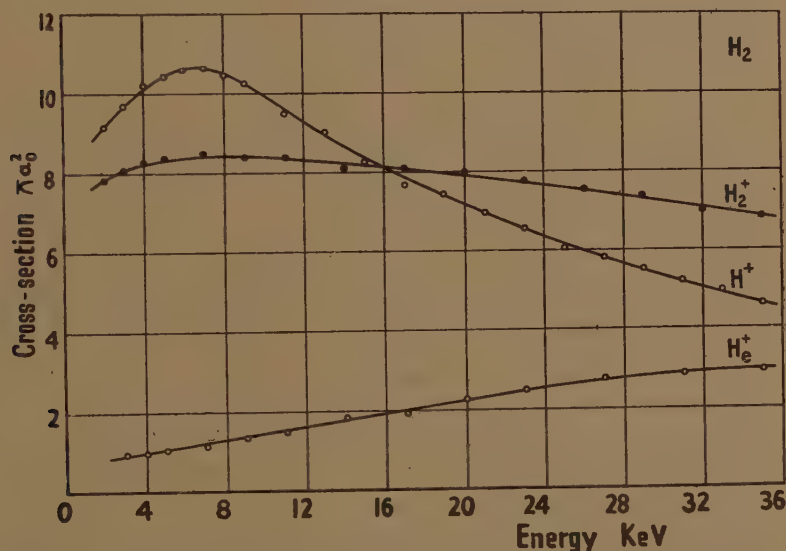
$$\sigma_i = \frac{I_-}{I_0} \cdot \frac{1}{NL}. \quad (6)$$

For H^+ , H_2^+ and He^+ ions in hydrogen the slopes of lines of the type shown in figs. 6 and 7 gave I_+/N and I_-/N , I_0 was measured as I_F , and the cross-sections were calculated directly. In order to obtain cross-sections as functions of the energy of the incident ion, it was necessary to repeat the measurements of I_F as a function of pressure for a series of energies in the range available.

For the same ions in helium, the charge exchange cross-sections were again found by subtracting the values of the slopes of the negative currents from the positive slopes. For the ionization cross-sections, curves of the type shown in fig. 10 were obtained for each energy of the incident ions in order to find the fraction of the electron current to P which was due to ionization. Then, for each energy the slope of the negative current to P (as a function of the pressure) was multiplied by the corresponding fraction when the cross-sections were calculated.

The results obtained for H^+ , H_2^+ , and He^+ ions in hydrogen and helium in the energy range 3–35 KeV. are shown in figs. 13, 14, 15 and 16. The unit of area used is πa_0^2 , the area of the first Bohr orbit in the hydrogen atom. ($\pi a_0^2 = 0.88 \times 10^{-16} \text{ cm.}^2$)

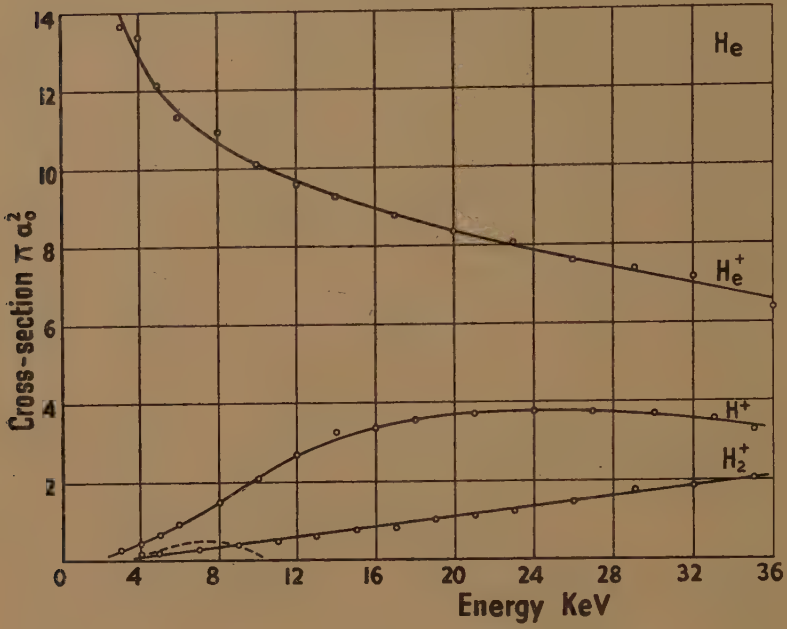
Fig. 13.



Charge exchange cross-sections for H^+ , H_2^+ , and He^+ ions in hydrogen.

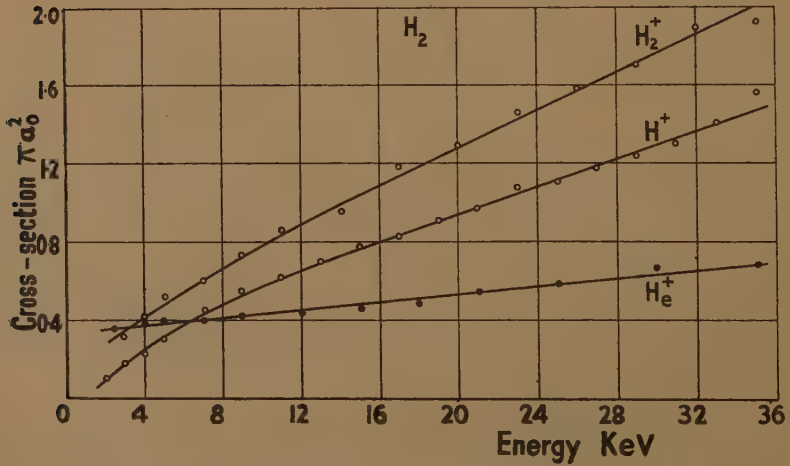
The probable error in the charge exchange cross-sections was determined by the uniformity of the electric field between P and B, the extent to which I_P saturated, and the errors in L , I_F , I_P , and N . Errors caused by non-uniformity of the field between P and B, incomplete saturation of I_P , and inaccuracy in the value of L were very small. The readings for I_F , I_P , and N could be taken with an accuracy of ± 2 –3 per cent and were repeatable. It thus remains to consider the absolute accuracies of I_F , I_P and N (*i. e.* accuracy of calibration). I_F was measured with a galvanometer, the calibration being correct to within ± 1 per cent. I_P depended on the accuracy of measurement of the high resistance in the grid circuit of the electrometer tetrode and this was ± 3 per cent. In the case of N , the accuracy depended on the calibration of a McLeod gauge followed by a comparison between the McLeod and an ionization gauge. N was estimated to be within ± 5 per cent. It follows that the

Fig. 14.



Charge exchange cross-sections for H^+ , H_2^+ , and He^+ ions in helium.

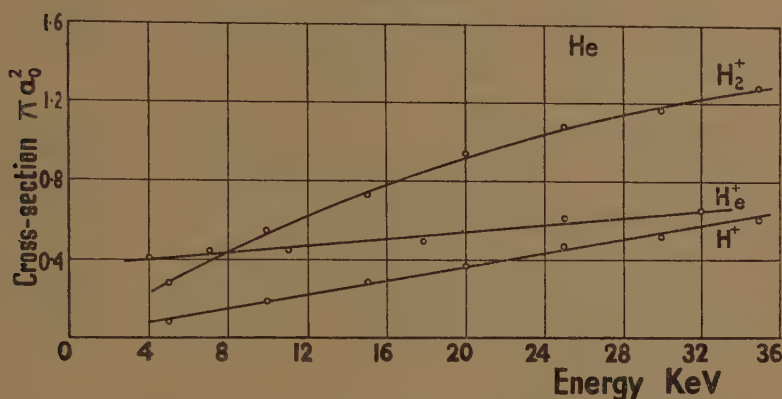
Fig. 15.



Ionization cross-sections for H^+ , H_2^+ , and He^+ ions in hydrogen.

probable error in the results was in the region of ± 10 per cent, most of this being due to systematic errors. For the ionization cross-sections

Fig. 16.



Ionization cross-sections for H^+ , H_2^+ , and He^+ ions in helium.

the currents were smaller and the probable error is estimated as ± 15 per cent for ions in hydrogen and ± 20 per cent for ions in helium.

V. DISCUSSION.

If an atom or molecule is regarded as a hard sphere the gas kinetic diameter can be calculated from the viscosity coefficient of the gas. Multiplying the square of this diameter by π gives the area through which the centre of an incident molecule must pass for a collision to occur with a stationary molecule. For hydrogen and helium the values of this area are approximately $20\pi a_0^2$ and $15\pi a_0^2$ respectively. These "gas kinetic collision cross-sections" may be compared with the results in figs. 13-16.

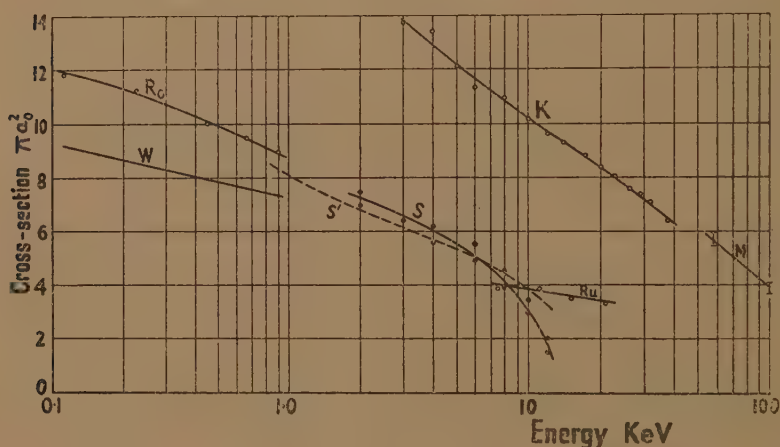
In the case of charge exchange cross-sections a number of results by other workers are available, particularly for He^+ ions in helium and H^+ ions in hydrogen. These are shown in figs. 17 and 18 compared with the present results.

The methods used by Goldmann (1931), Rostagni (1939), and Wolf (1936, 1937 b) are similar in principle to the present method although they differ considerably in detail. Rudnick (1931), Bartels (1932), and Meyer (1937) used various forms of an equilibrium method originally suggested by Wien (1912). This method depends on the fact that a fast neutral particle, formed by charge exchange, may lose an electron in a further collision (without suffering appreciable change in kinetic energy or direction) and thus become a positive ion again. The ratio of the number of neutrals to charged particles in a beam is measured; this ratio depends on the gas pressure but it reaches an equilibrium value for pressures greater than about 10^{-2} mm./Hg. and path lengths of 10-20 cm. From the equilibrium value, together with a non-equilibrium value at a known pressure and path length, the cross-sections for charge

exchange and for the reverse process can be calculated. From wave-mechanical considerations, Smith (1934b) calculated charge exchange cross-sections for He^+ ions in helium; he also made a number of measurements for various ions and gases. The method used was to measure the beam current for three different path lengths in the gas at various pressures, using a Faraday cage with a large aperture.

For He^+ ions in helium (fig. 17) there is fair agreement between the present results and those of Meyer. On the other hand the results of Rudnick, Wolf, Rostagni, and Smith (theoretical and experimental) are very much lower and all agree to some extent. It is difficult to understand why Smith's experimental curve falls to such low values for energies greater than 10 KeV.; his results also fall to very low values in fig. 18 where charge exchange cross-sections for H^+ ions in hydrogen

Fig. 17.



Charge exchange cross-sections for He^+ ions in helium compared with results of other workers. K, present results. W, Wolf. Ro, Rostagni. M, Meyer. Ru, Rudnick. S, Smith (experimental). S', Smith (theoretical).

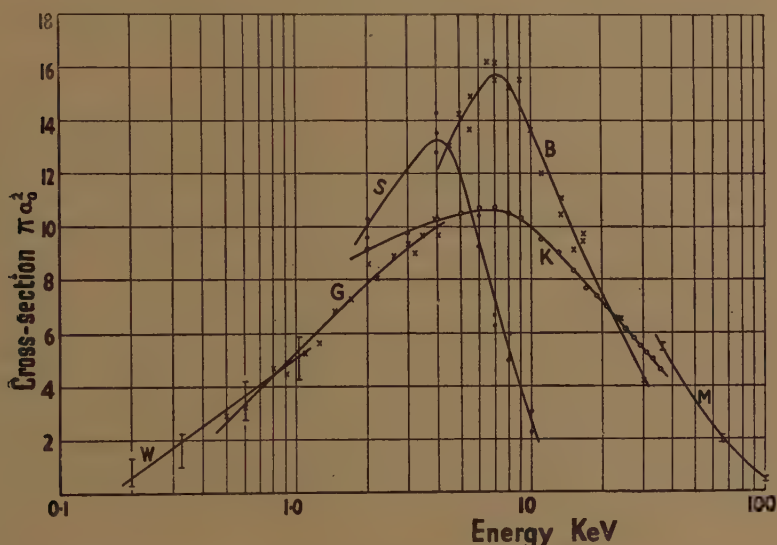
are shown. The other curves in fig. 18 agree moderately well, although Bartels results reach a rather large maximum at 7.1 KeV. A small extension of Smith's curve to 13 or 14 KeV. suggests cross-sections which would only be a few per cent of the other results; this is difficult to reconcile with Meyer's measurements up to 200 KeV. Smith also measured the cross-section for H^+ ions in helium and his results are shown as the small dashed curve in fig. 14. In view of the general disagreement between Smith's results and those of other workers, it is hard to understand his agreement with theory and with Wolf and Rostagni in the particular case of He^+ ions in helium. His experiments were carried out in a glass bulb and it is possible that secondary emission from the walls (by charge exchange ions or metastable atoms) was upsetting his measurements.

The few other charge exchange cross-sections available include those of Meyer (1937) for H^+ ions in helium and He^+ ions in hydrogen for

energies greater than 40 KeV. Both curves agree fairly well with the present ones.

Very few results are available for ionization cross-sections. In the case of He^+ ions in helium Wolf (1935) and Rostagni (1939) have made measurements for energies less than 1 KeV. and their results agree with

Fig. 18.



Charge exchange cross-sections for H^+ ions in hydrogen compared with results of other workers. K, present results. W, Wolf. G, Goldmann. M, Meyer. B, Bartels. S, Smith.

the present curve if it is extrapolated. For 1 KeV. H_2^+ ions in hydrogen Wolf (1937a) was unable to detect ionization and Goldmann (1931) could not detect it for 4 KeV. H^+ ions in hydrogen.

ACKNOWLEDGMENTS.

The author is indebted to Mr. E. S. Shire and to Professor H. S. W. Massey for helpful advice, and to the Department of Scientific and Industrial Research for a maintenance Allowance.

REFERENCES.

- BARTELS, H., 1932, *Ann. Phys., Lpz.*, **13**, 373.
 GOLDMANN, F., 1931, *Ann. Phys., Lpz.*, **10**, 460.
 MEYER, H., 1937, *Ann. Phys., Lpz.*, **30**, 635.
 PENICK, D. B., 1935, *Rev. Sci. Instrum.*, **6**, 115.
 ROSTAGNI, A., 1939, *Nuovo. Cim.*, **15**, 117.
 RUDNICK, P., 1931, *Phys. Rev.*, **38**, 1342.
 SHERWIN, C. W., 1940, *Phys. Rev.*, **57**, 814.
 SMITH, R. A., 1934a, *Proc. Camb. Phil. Soc.*, **30**, 514; 1934b, *Ph.D. Thesis* (Cambridge University).
 WIEN, W., 1912, *Ann. Phys., Lpz.*, **39**, 519.
 WOLF, F., 1935, *Ann. Phys., Lpz.*, **23**, 627; 1936, *Ibid.*, **27**, 543; 1937a, *Ibid.*, **29**, 33; 1937b, *Ibid.*, **30**, 313.

XXXIV. Notes on the Molecular Orbital Treatment of the Hydrogen Molecule.

By Prof. C. A. COULSON and Miss I. FISCHER,
Wheatstone Physics Department, King's College, London*.

[Received January 5, 1949.]

§ 1. INTRODUCTION.

THE molecular orbital (m.o.) method has been used regularly for many years (Coulson 1947), but relatively little is known about its validity except on semi-empirical grounds. The purpose of these notes is to investigate the fundamentals of the m.o. method in its LCAO form (linear combination of atomic orbitals) somewhat more fully than before. For this study we have chosen the hydrogen molecule. By virtue of its simplicity this molecule allows us to make calculations which would be prohibitively difficult for larger systems. Now both the ground state and several excited states of H_2 have been successfully treated by other authors. Our object is not to achieve a better agreement with experiment than that already obtained, but to examine the possibilities and limitations of the m.o. method in its simplest form, and including certain refinements. Some of our considerations have been published earlier, though in a less completed form (see a paper by Coulson (1937) and a review by Van Vleck and Sherman (1935)), but we believe that the rest, particularly the detailed numerical values, are here made available for the first time.

If we confine ourselves to molecular orbitals composed out of atomic $1s$ functions ψ_a and ψ_b of the atoms A and B, the allowed m.o. are ϕ and χ , where

$$\phi = (\psi_a + \psi_b) / \{2(1 + S)\}^{\frac{1}{2}}, \quad (1)$$

$$\chi = (\psi_a - \psi_b) / \{2(1 - S)\}^{\frac{1}{2}}, \quad (2)$$

and

$$S = \text{overlap integral} = \int \psi_a(1) \psi_b(1) d\tau. \quad (3)$$

If ψ_a and ψ_b are exactly the same as the atomic orbitals for an isolated hydrogen atom (perturbation method of Coulson (1937)), then, in atomic units,

$$\psi_a(1) = \sqrt{\frac{1}{\pi}} e^{-r_{a1}}. \quad (4)$$

But if we include the possibility of a partial screening (variation method of Coulson (1937)), we write

$$\psi_a(1) = \sqrt{\frac{c^3}{\pi}} e^{-cr_{a1}}, \quad (5)$$

* Communicated by the Authors.

where c is a constant which varies with the internuclear distance R and which may be calculated in the standard manner. We shall be concerned with the inter-relations of the various molecular states that arise from different possible occupations of the orbitals ϕ and χ . No details of the manipulative part of our calculations need be given, for they follow completely conventional lines, once the appropriate wave functions for the molecule have been set up.

§ 2. FAILURE AT LARGE INTERNUCLEAR DISTANCES IN THE GROUND STATE.

It is well recognized that at large internuclear distances R the m.o. method fails badly because it takes insufficient account of the electron repulsion. This repulsion ensures that when the molecule dissociates one electron is found on each nucleus. But the conventional m.o. description of the ground state, viz.

$$\Psi = \phi(1)\phi(2) \times \text{spin term}, \quad (6)$$

divides each electron equally between A and B. It is important to know at what value of R the wave function (6) falls seriously in error. This may be done as follows.

We may imagine that as R increases, one of the two m.o. tends to concentrate more round atom A, and the other around atom B. This could be achieved by replacing the m.o. $\phi(1)$ and $\phi(2)$ in (6) by two different orbitals

$$\psi_a + \lambda\psi_b, \quad \psi_b + \lambda\psi_a, \quad (7)$$

where λ is a parameter not necessarily equal to 1. Assigning one electron to each of these orbitals we describe the molecular system by the wave function

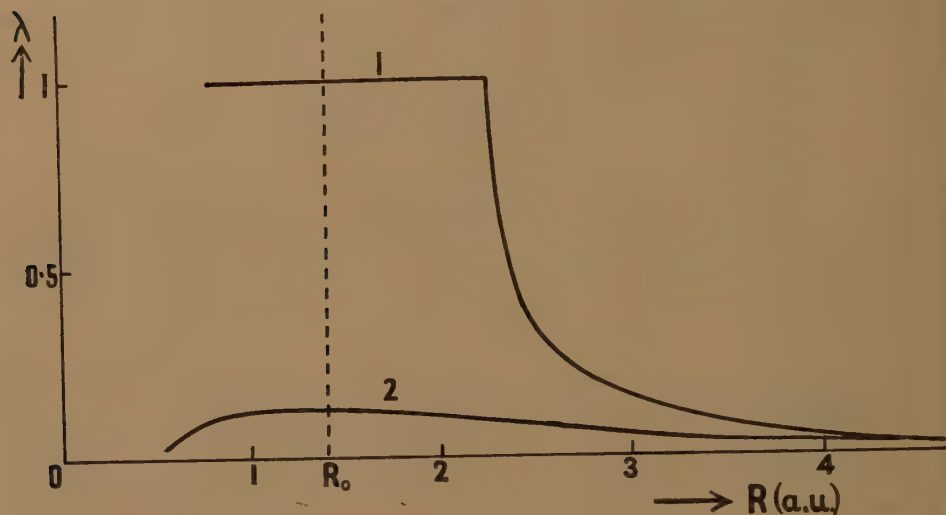
$$\{\psi_a(1) + \lambda\psi_b(1)\}\{\psi_b(2) + \lambda\psi_a(2)\}. \quad (8)$$

Now, the parameter λ has to be determined by a variation method. For those values of R for which λ turns out to be 1, we can say that the ordinary m.o. is a valid type of approximation; but for those values for which λ differs considerably from 1, we interpret the significance of the wave function (8) to be that the repulsion between the electrons, tending to separate them on to different nuclei, is stronger than the additional attraction which arises when both electrons are under the influence of two nuclei.

We have chosen the appropriate value of λ for each R by minimizing the "energy" function $\int \Psi^* H \Psi d\tau / \int \Psi^* \Psi d\tau$, where H is the Hamiltonian of the hydrogen molecule. For this purpose we used the values of the exponent c in (5) which had previously been calculated for the true m.o. wave function (6) by Coulson (1937). Fig. 1 (curve 1) shows the value of λ plotted against R . Up to $R = 2.27$ a.u. (*i.e.* $R = 1.6 R_0$, where R_0 is the equilibrium distance in H_2) λ is exactly equal to 1. But when $R > 1.6 R_0$, λ rapidly falls almost to zero, indicating a complete breakdown in the simple m.o. theory.

It is very probable that these conclusions hold for other molecules. In that case our calculation shows very clearly the dangers inherent in too naive an application of m.o. theory to interactions across large

Fig. 1.



Values of the parameter λ in unsymmetrical m.o.'s (7) giving energy minimum with: curve 1: an asymmetrical wave function (8); curve 2: a wave function including exchange (9).

distances, as, for example, in trying to follow the complete course of a chemical reaction (Coulson and Dewar 1947).

§ 3. OVER-EMPHASIS OF IONIC TERMS.

A second criticism of m.o. wave functions such as (6) is that even at the equilibrium distance they overemphasize ionic terms. We may discuss this also with the aid of the unsymmetrical molecular orbitals (7). If we suppose that those are the occupied m.o., then the true wave function, in which antisymmetrization is employed, is, apart from a normalization factor and a spin term:

$$\Psi' = \{\psi_a(1) + \lambda\psi_b(1)\}\{\psi_b(2) + \lambda\psi_a(2)\} + \{\psi_b(1) + \lambda\psi_a(1)\}\{\psi_a(2) + \lambda\psi_b(2)\}. \quad (9)$$

If we write this in the equivalent form

$$\Psi' = (1 + \lambda^2)\{\psi_a(1)\psi_b(2) + \psi_b(1)\psi_a(2)\} + 2\lambda\{\psi_a(1)\psi_a(2) + \psi_b(1)\psi_b(2)\}, \quad (10)$$

we recognize it as the wave function used by Weinbaum (1933), who actually used

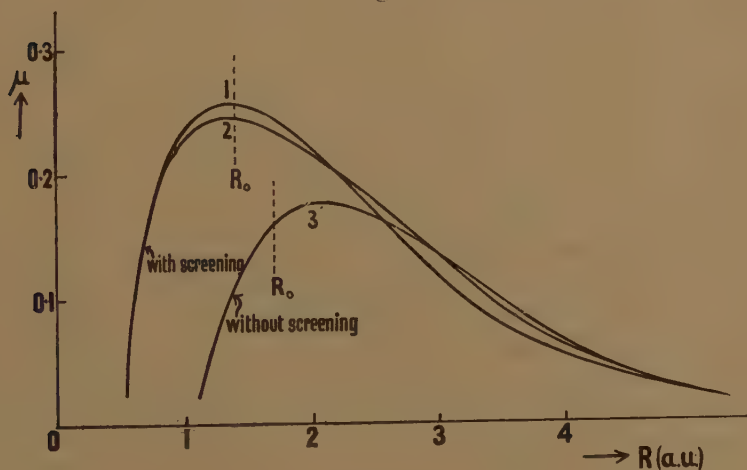
$$\Psi' = \{\psi_a(1)\psi_b(2) + \psi_b(1)\psi_a(2)\} + \mu\{\psi_a(1)\psi_a(2) + \psi_b(1)\psi_b(2)\}. \quad (11)$$

The expressions in brackets are the simple Heitler-London covalent

wave function and the pure ionic wave function; μ shows the degree of mixing between them. It is interesting—and, to us, new—that the conventional covalent-ionic resonance can be described equally well by the introduction of asymmetrical molecular orbitals, as in (9). The significance of this equivalence is that if we wish to refine the m.o. theory, while still retaining the idea of a molecular orbital, one possibility is to abandon the m.o. ϕ in equation (1) in favour of the pair of functions (7). If it turns out that with the wave function (9), the value of λ is not equal to 1, then the best description of the molecule in terms of a pair of m.o. composed linearly of ψ_a and ψ_b , requires the two orbitals to be different.

Values of μ which minimize the energy function for (11) are shown in fig. 2 in terms of the internuclear distance R . The three curves shown relate to different choices of the screening constant c . In curve 1 we used

Fig 2



Contribution μ of ionic terms calculated with best screening constant of:— curve 1: a molecular orbital wave function; curve 2: a Heitler-London wave function; curve 3: atomic orbitals.

The dotted lines show the values of the equilibrium distance R_0 calculated with these wave functions.

the values calculated for a pure m.o. wave function by Coulson (1937); in curve 2 we used the values appropriate to a pure Heitler-London wave function; and in curve 3 we used the atomic wave functions (perturbation method) for which $c=1$. Fig. 2 shows that provided some allowance is made for screening, the value of μ is not particularly sensitive to c ; but if no allowance whatever is made, values of μ in error by as much as 70 per cent, may occur. As we should expect, the ionic contribution to the ionic-covalent hybrid decreases with increasing R .

A comparison of (10) and (11) allows us to convert our μ -values into λ -values. The conversion appropriate to the second curve in fig. 2 is shown as curve 2 in fig. 1, where it may be contrasted with the curve

already drawn on the basis of wave function (8). This shows very clearly how grossly the naive m.o. theory with $\lambda=1$, exaggerates the importance of ionic terms, and how, though in a smaller degree, the asymmetrical function (8) also does so. As we might have anticipated, antisymmetrization, or exchange, has the effect of reducing the two-centre character of the molecular orbitals, *i. e.*, of making them more asymmetrical.

§ 4. EXCITED STATES ; CONFIGURATIONAL INTERACTION.

There is a third point of view from which we may question the validity of the naive m.o. treatment. The idea is thoroughly discussed in a qualitative way by Hund (1932), but the only published numerical results that we have found, are incomplete and take no account of any screening (Slater 1930, Mulliken 1932). If we start with the two m.o. ϕ and χ of (1) and (2), we can form four molecular states $\Psi_1 \dots \Psi_4$. Neglecting spin terms, these are

$$\left. \begin{aligned} \Psi_1 &= \phi(1)\phi(2), & {}^1\Sigma_g \\ \Psi_2 &= \phi(1)\chi(2) - \chi(1)\phi(2), & {}^3\Sigma_u \\ \Psi_3 &= \phi(1)\chi(2) + \chi(1)\phi(2), & {}^1\Sigma_u \\ \Psi_4 &= \chi(1)\chi(2). & {}^1\Sigma_g \end{aligned} \right\} \dots \dots \dots (12)$$

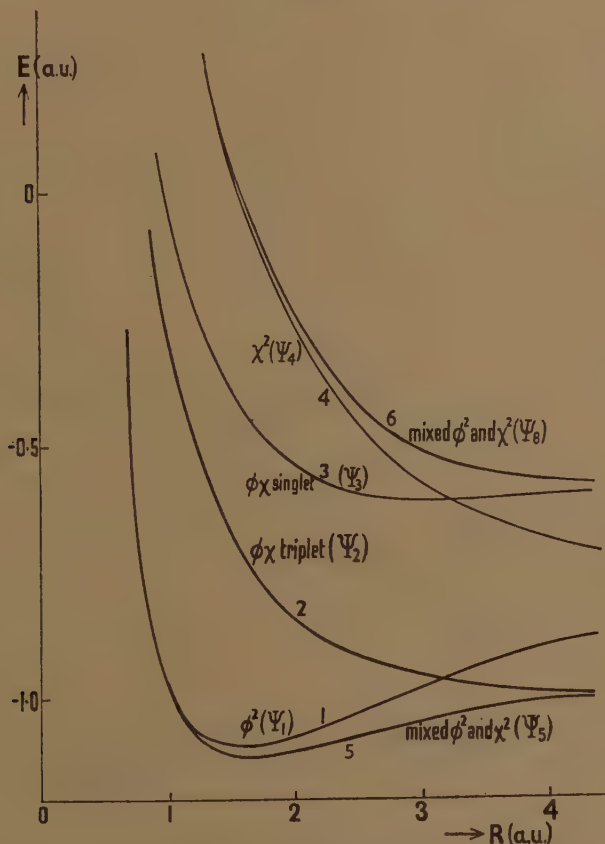
These may be said to be the wave functions associated with the "pure configurations" ϕ^2 , $\phi\chi$ and χ^2 . If our molecular orbital description of the molecule is sound, these configurations should be practically non-interacting with each other: and also the energy of the singlet function Ψ_3 , which is an ionic function and represents the main part of the experimentally found B state, and which corresponds to excitation of one electron in a process $\phi^2 \rightarrow \phi\chi$, should lie between that of ϕ^2 , in which there is no excitation, and that of χ^2 in which two electrons are excited. Now Craig (1949) has given reasons for believing that among the π -electrons of ethylene, where states analogous to those of (12) occur, the energy of the configuration χ^2 is either just below, or closely similar to, the energy of $\phi\chi$. This makes it desirable to make similar calculations for H_2 , where the approximations in calculating the energy are less severe than for ethylene. We have therefore calculated—and show in fig. 3—the energies appropriate to the four pure configurations $\Psi_1 \dots \Psi_4$. In these calculations we used the atomic orbitals (4) rather than (5) because that seemed the best compromise between values $c > 1$ for the ϕ orbitals and $c < 1$ for the χ orbitals (Coulson 1937). Fortunately the main character of these curves is still maintained if, instead, we use the values of c found most suitable for the ϕ orbitals. It will be noticed that at fairly large internuclear distances ($R > 3.2$ a.u.) the energies of the singlets $\phi\chi$ and χ^2 are in an inverted order. The interpretation of this is that our pure configurations interact with each other. So far as the configurations listed in (12) are concerned, this only involves Ψ_1 and Ψ_4 , for these are

the only ones with equivalent symmetry. We ought, therefore, to replace Ψ_1 and Ψ_4 by two "mixed configurational" wave functions

$$\left. \begin{aligned} \Psi_5 &= \Psi_1 + \nu \Psi_4, \\ \Psi_8 &= \Psi_1 + \nu' \Psi_4, \end{aligned} \right\}, \quad \dots \quad (13)$$

where ν and ν' are two constants such that $\nu\nu' = -1$. Explicit expansion of Ψ_5 shows that it is of exactly the same form as (10), so that we may use

Fig. 3.

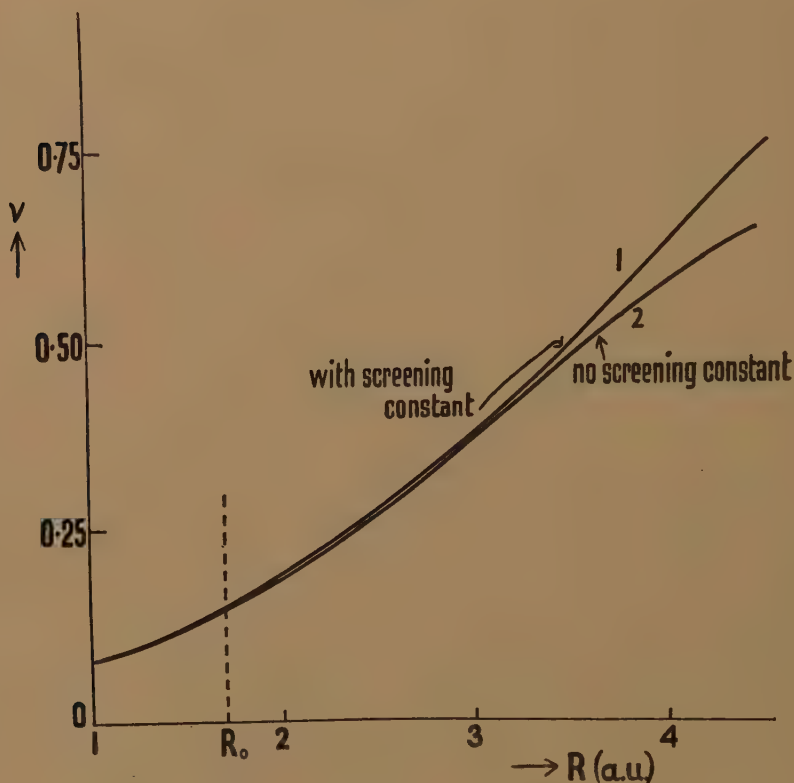


Energy curves for the hydrogen molecule. Curves 1 4 correspond to the "pure configurations" Ψ_1 Ψ_4 of (12), curves 5 and 6 to Ψ_5 and Ψ_8 (13) which account for configurational interaction between Ψ_1 and Ψ_4 .

our calculations with this wave function to infer the value of ν corresponding to any chosen R , without further labour. The value $\nu \simeq 1/8$ at $R = R_0$ has already been given by Slater (1930) and used by Mulliken (1932). Our fig. 4 shows that for large distances ν increases, indicating an increasing configurational interaction, until, at $R \rightarrow \infty$, we find $\nu \rightarrow 1$. The figure also shows that the degree of interaction is almost independent of the choice of screening constant.

This configurational interaction lowers the energy of the state Ψ_I and raises that of Ψ_4 , so that, as fig. 3 shows, the unsatisfactory crossing-over of the energies $\phi\chi$ and χ^2 no longer takes place, even at large distances. For reasonably small R the configurational representation in terms of ϕ and χ is valid—though there is always some degree of interaction between configurations—but for R larger than about $1.5 R_0$ this interaction is so large that the ϕ, χ representation becomes quite invalid.

Fig. 4.



Configurational interaction constant ν calculated with:—curve 1: best screening constant of molecular orbital wave function; curve 2: atomic orbitals.

As already pointed out, Ψ_5 and Ψ_8 are of the same form as (10) and are thus equivalent to the use of asymmetrical molecular orbitals. We may make the parallel between conventional descriptions of the wave functions for H_2 and our new description a little clearer if we label the two asymmetrical m.o.'s (7) A_+ and B_+ , and we introduce two related orbitals A_- , B_- , where

$$\left. \begin{aligned} A_{\pm} &= \psi_a \pm \lambda \psi_b, \\ B_{\pm} &= \psi_b \pm \lambda \psi_a. \end{aligned} \right\} \dots \dots \dots (14)$$

It is easily verified that the "mixed configurations" which result from interaction among the group (12) are represented by the four wave functions

$$\left. \begin{aligned} \Psi_5 &= A_+(1)B_+(2) + B_+(1)A_+(2), & 1\Sigma_g \\ \Psi_6 &= \{A_+(1)B_-(2) - B_-(1)A_+(2)\} - \{A_-(1)B_+(2) - B_+(1)A_-(2)\}, & 3\Sigma_u \\ \Psi_7 &= \{A_+(1)B_-(2) + B_-(1)A_+(2)\} - \{A_-(1)B_+(2) + B_+(1)A_-(2)\}, & 1\Sigma_u \\ \Psi_8 &= A_-(1)B_-(2) + B_-(1)A_-(2). & 1\Sigma_g \end{aligned} \right\} \quad (15)$$

The energies appropriate to these four wave functions are given by curves 5, 2, 3 and 6 respectively of fig. 3.

Now the Σ_u states in (15) are identical with those of (12). But the Σ_g states could, if we wished, be regarded as arising from pure configurations A_+B_+ and A_-B_- . Indeed, to this degree of approximation, we can avoid the annoying configurational interaction between Ψ_1 and Ψ_4 if we replace the symmetrical molecular orbitals (1) and (2) by the asymmetrical ones. (14), giving the non-interacting configurations Ψ_5 and Ψ_8 . If we are to use a configurational representation at all, it must be in terms of these latter functions. In this way we get further light on the discussion in § 3.

In any complete treatment of this molecule, we ought also to include further interaction with other configurations such as those built on the $2s$ or $2p$ atomic orbitals. This would depress all the levels slightly, but it is unlikely that their sequence would be changed.

In conclusion we wish to point out that the configurational interaction to which we have drawn attention is likely to be particularly significant when the energies of the levels are closer together than in H_2 . We may therefore anticipate that some of the difficulties already found in interpreting the u.v. spectra of large condensed molecules arise from its neglect. Work in progress in this department suggests that this is indeed the case (Miss J. Jacobs, unpublished calculations).

One of us (I. F.) would like to acknowledge a grant from the Swedish State Council for Scientific Research—(Statens Naturvetenskapliga Forskningsrad)—which has made possible this work.

REFERENCES.

- COULSON, C. A., 1937, *Trans. Faraday Society*, **33**, 1479; 1947, *Quart Reviews*, **1**, 144.
 CRAIG, D. P. In course of publication.
 HUND, F., 1932, *Zeit. Physik*, **73**, 1,
 JACOBS, Miss J. Unpublished calculations.
 MULLIKEN, R. S., 1932, *Phys. Rev.*, **41**, 49, especially pp. 68–70.
 SLATER, J. C., 1930, *Phys. Rev.*, **35**, 509, especially pp. 514–5.
 VAN VLECK, J. H., and SHERMAN, A., 1935, *Rev. Modern Physics*, **7**, 167.
 WEINBAUM, S., 1933, *J. Chem. Phys.*, **1**, 593.

XXXV. *Angular Distribution, Frequency and Absorption of Slow Single Cosmic Ray Protons.*

By S. LATTIMORE,
Imperial College, London *.

[Received January 3, 1949.]

§ 1. INTRODUCTION.

IN special photographic plates exposed to Cosmic Rays, there are observed tracks, due to protons, other than those forming "stars". Previous investigations of the frequency of occurrence of these tracks led to the conclusion that these tracks were produced by the protons from the "stars" in the surrounding material (Perkins 1947a, Widhalm 1940). In the present paper, the results of an investigation into the angular distribution, frequency and absorption of these protons are given.

PART A. *Angular Distribution and Frequency.*

1. Since we observe protons emitted from the stars in the emulsion, we would expect at least some of the observed single protons to be due to stars in the surrounding material. This effect has been estimated on the assumption that the protons emitted from these stars have the same frequency, energy and angular distributions as those emitted from stars in the emulsion.

The mean number of protons emitted per star in the emulsion is found to be about three (Bernadini, Cortini and Manfredini 1948). It has been shown (Perkins 1947b) that the energy distribution of the protons can be fitted with an evaporation formula, of the form

$$N(E) dE = \frac{E-V}{T^2} e^{-\frac{E-V}{T}} dE, \quad (1)$$

where $N(E) dE$ is the probability of an emitted proton having energy between E and $E+dE$, V is the height, in MeV., of the potential barrier of the nucleus forming the star, and T is a constant, with the dimensions of an energy, called the temperature and measured in MeV. We find that the experimental results are best fitted with a value of $V=3$ MeV. and T equal approximately to 4 MeV. At energies above about 25 MeV., however, more protons are observed than would be expected from this formula. In these calculations, this high energy "tail" has been ignored. The protons are found to be emitted isotropically from the stars. The number of stars at 3600 m. is taken to be 3.6 gm./day.

* Communicated by Professor G. P. Thomson.

§ 2. THEORETICAL CALCULATIONS.

(a) Horizontal Plate.

Let θ be the angle between the protons and the vertical; $P(\theta)$ be the differential angular distribution of single protons, as observed per cm.² in the plate; $Q(l, \theta)$ be the differential angular distribution of the protons with range greater than l formed/cc. in the surroundings (*i. e.* $Q(l, \theta) d\Omega dv$ = number of protons with range greater than l emitted into solid angle $d\Omega$, at angle θ to the vertical, from volume dv).

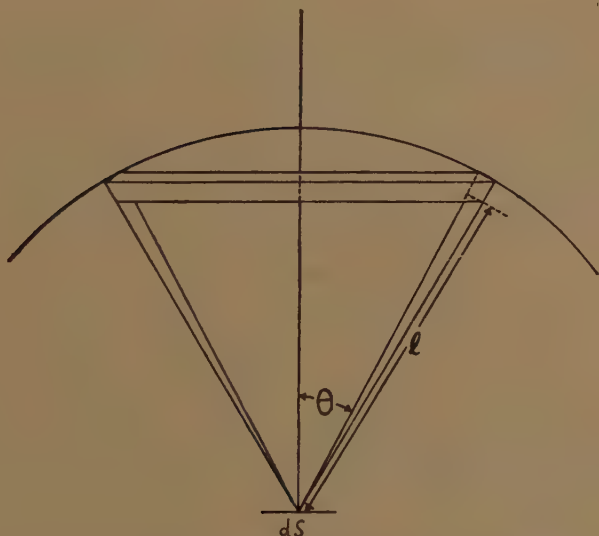
Then

$$P(\theta) d\theta = \int_0^\infty 2\pi l^2 \sin \theta d\theta \frac{\cos \theta}{l^2} \{Q(l, \theta) + Q(l, \theta + \pi)\} dl.$$

$$\therefore P(\theta) = \pi \sin 2\theta \int_0^\infty \{Q(l, \theta) + Q(l, \theta + \pi)\} dl \quad . \quad . \quad . \quad (2)$$

(see fig. 1).

Fig. 1.



$Q(l, \theta) + Q(l, \theta + \pi)$ is used instead of $Q(l, \theta)$ since, in general, the direction of the protons cannot be found.

For an isotropic distribution, $Q(l, \theta) = Q(l)$ is independent of θ , and therefore

$$\frac{P(\theta)}{\sin 2\theta} = 2\pi \int_0^\infty Q(l) dl. \quad . \quad . \quad . \quad (3)$$

Let N be the total number of protons/cm.² of plate. Then

$$\begin{aligned} N &= \int_0^{\pi/2} P(\theta) d\theta \\ &= \int_0^{\pi/2} \int_0^\infty \pi \sin 2\theta \{Q(l, \theta) + Q(l, \theta + \pi)\} dl d\theta. \quad . \quad . \quad . \quad (4) \end{aligned}$$

∴ for an isotropic distribution this reduces to

$$N = 2\pi \int_0^\infty Q(l) dl, \quad (5)$$

and from (3) and (5)

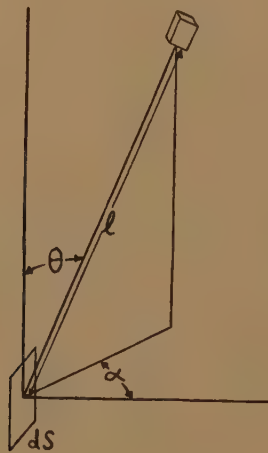
$$\frac{P(\theta)}{\sin 2\theta} = N. \quad (6)$$

We would expect relation (6) to hold if all the protons were due to stars.

(b) *Vertical Plate.*

Let θ be the angle between the protons and the vertical, $P'(\theta)$ be the differential angular distribution of single protons as observed/cm.² in the plate.

Fig. 2.



$$P'(\theta) d\theta = 4 \int_0^{\pi/2} \int_0^\infty \{Q(l, \theta) + Q(l, \theta + \pi)\} \sin \theta d\theta \cos \alpha \sin \theta dl dx.$$

$$\therefore P'(\theta) = 4 \sin^2 \theta \int_0^\infty \{Q(l, \theta) + Q(l, \theta + \pi)\} dl \quad (7)$$

(see fig. 2).

∴ From eqns. (2) and (7)

$$\frac{P'(\theta)}{P(\theta)} = \frac{2}{\pi} \tan \theta. \quad (8)$$

∴ if $P(\theta)$ is known, $P'(\theta)$ can be found.

§ 3. EXPERIMENTAL RESULTS.

The plates used were Ilford 100 μ C.2 emulsion on glass 1 mm. thick, and these were exposed, at 3600 meters above sea level. The plates

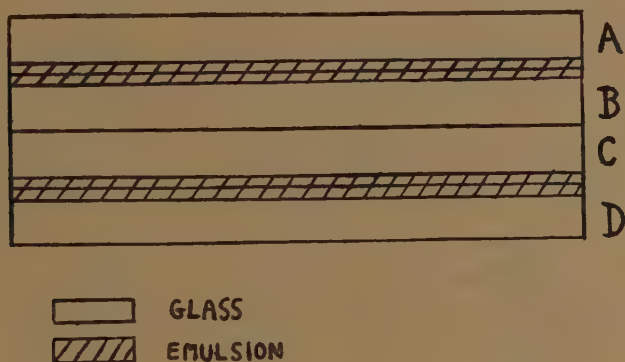
were packed as shown in fig. 3. They were surrounded with black wrapping paper, and a cardboard box. Table I. shows the observed frequency of tracks/cm.²/day for horizontal plates.

TABLE I.

Plate	Tracks/cm. ² /day
A	16.4 ± 0.8
C	16.6 ± 0.8
Mean value	16.5 ± 0.6

Approximately 4 per cent of these tracks stop in the emulsion; a further 2 per cent start in the emulsion, of which approximately half stop in the emulsion as well. The contribution of single tracks from the stars in the surrounding medium has been estimated at 1.5 cm.²/day using eqns. (1) and (5) and the distributions given in § 1.

Fig. 3.



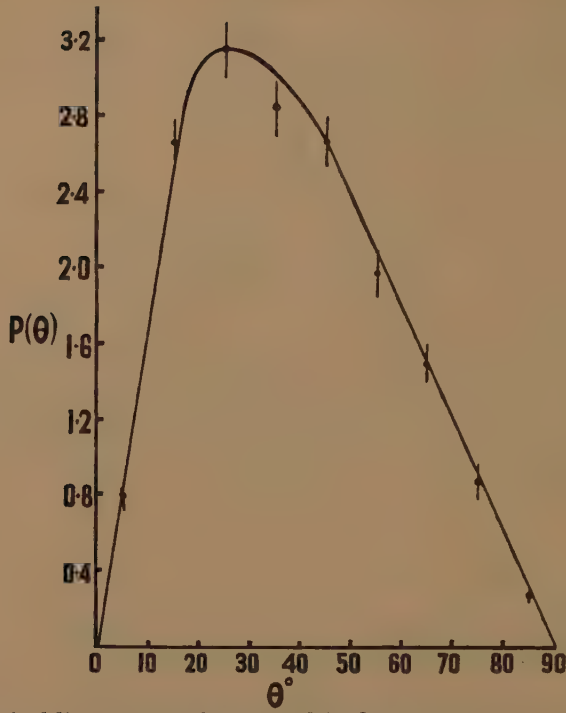
Thus the observed number of single tracks is very much greater than would be expected from the stars in the surrounding material.

In fig. 4*a* is shown the observed angular distribution $P(\theta)$ for a horizontal plate. Fig. 4*b* shows $P(\theta)/\sin 2\theta$ as ordinates. This has a maximum at about $\theta=0$, and drops to a finite value at $\theta=\pi/2$. From this finite value, we can calculate the maximum number of protons which could have an isotropic distribution, since for an isotropic distribution

$$\frac{P(\theta)}{\sin 2\theta} = N,$$

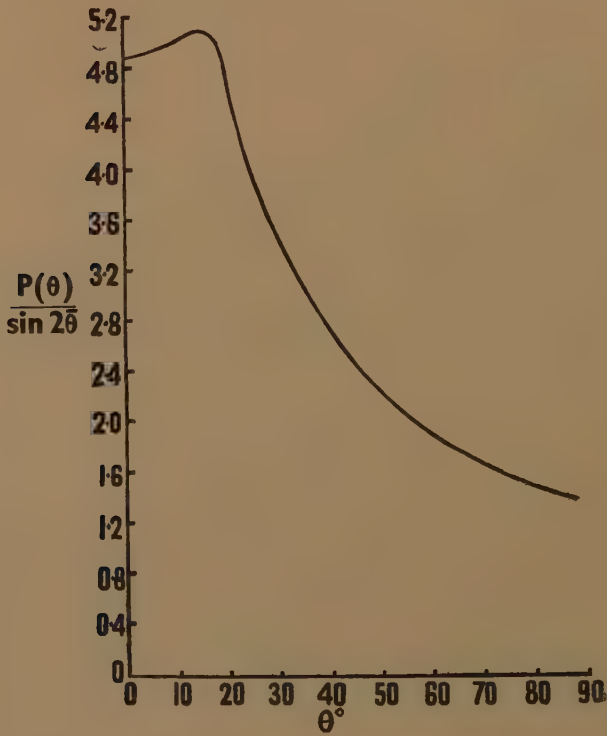
the total number of protons/cm.² (eqn. (6)) and for no distribution can $P(\theta)/\sin 2\theta$ be negative. This value is 1.4 cm.²/day in close agreement with the number to be expected from the stars. This suggests that the observed distribution should be divided into two parts, one being isotropic and due to the stars, the second showing a maximum at about $\theta=0$. It would seem likely that the protons in the second group were coming downwards, *i. e.* $Q(1, \theta+\pi)=0$.

Fig. 4 *a*.



The vertical lines in this figure, and in fig. 3 represent the probable error of the experimental results.

Fig. 4 *b*.



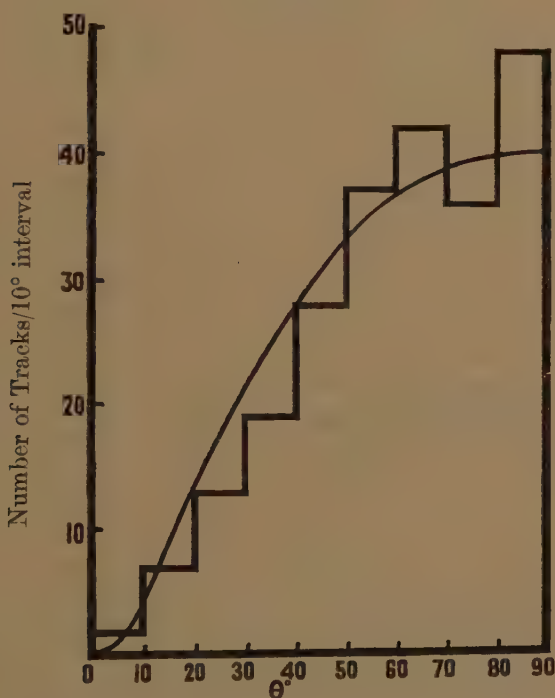
The distribution in fig. 4a has been checked by the following methods.

1. Using eqn. (8), $P'(\theta)$, the angular distribution for a vertical plate, can be calculated. This is shown, multiplied by a normalizing factor, in fig. 5, the histogram representing the experimental results.

2. The ratio of the number of tracks/cm.²/day on a horizontal plate to the number/cm.²/day on a vertical plate was calculated. This value is 1.32. The experimental value is 1.35 ± 0.15 .

The number of single tracks/cm.²/day was also measured using 200 μ Kodak NT/2a emulsions on glass. These plates have greater sensitivity than the Ilford C.2. The number of tracks/cm.²/day was found to be about 50 per cent greater than on the C.2 plates.

Fig. 5.



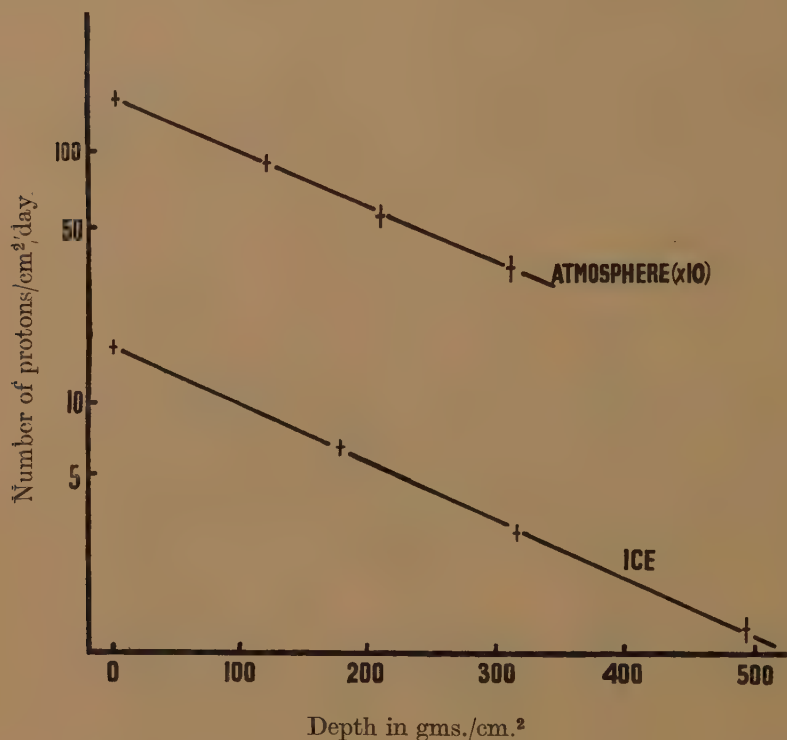
PART B. Absorption of the Protons in the Atmosphere and in Ice.

100 μ C.2 emulsions, packed as in fig. 3, were used.

1. *Atmosphere.* Plates were exposed at various altitudes in Switzerland. Unfortunately, the thicknesses of the roofs above the plates were unknown. It is assumed that all the plates had the same thickness of roof. The results are shown in fig. 6. As can be seen, we can write $N = N_0 e^{-\frac{h}{R}}$, where N_0 and R are constants, and h is the depth in the atmosphere, measured from 3600 m., in gms./cm.². The value of R found by the method of least squares is 190 gm./cm.²

2. *Ice.* Plates were exposed in a crevasse at the Jungfraujoeh at 3600 m., and the absorption curve obtained (fig. 6). Again we can write $N=N_0e^{-\frac{h}{R}}$, h being measured in gm./cm.², with $R=196$ gm./cm.²

Fig. 6.



CONCLUSIONS.

The slow protons cannot all be produced by stars similar to those observed in the emulsion since the number observed/cm.²/day is too great.

Hydrogen knock-ons can be excluded for the following reason. The number of protons observed under 3 mm. of glass is equal to the number observed under 1 mm. (see Table I.). Since glass contains practically no hydrogen, the knock-ons would have to have a range large compared to 3 mm. of glass. We would therefore expect a large increase in the number of protons observed when the plates were surrounded by ice instead of air (which contains very little hydrogen). This increase was not found.

Bagge (1943) has suggested that the excess of protons is due to one pronged stars, with an energy distribution similar to that of the many-pronged stars. We can calculate the number of such one-pronged stars in the following way.

Let n be the number of multi-pronged stars/cc. Then, since the mean number of protons emitted/multi-pronged star is three, the number of protons formed/cc. by multi-pronged stars $= 3n$. This number of protons will only explain about 10 per cent of the observed single protons. Therefore, the total number of protons emitted/cc. must be $30n$, and the number formed by one-pronged stars is therefore equal to $27n$, *i. e.* there are 27 one-pronged stars to every multi-pronged star. Since, on a $100\ \mu$ plate 0.15 multi-pronged stars are observed/cm.²/day, there should be approximately 4.5 one-pronged stars/cm.²/day. This is about 30 per cent of the observed number of protons.

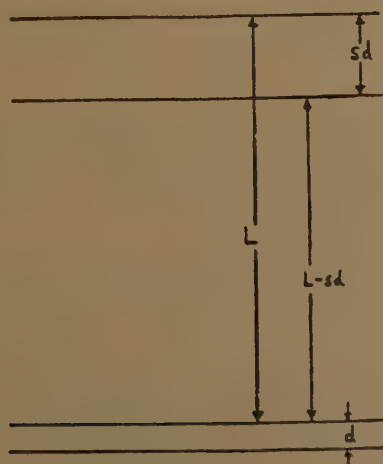
We actually observe that only about 3 per cent of the single tracks start in the emulsion, of which half also end in it. It is probable that most of this 3 per cent does represent the true one-pronged stars.

It can also be shown that the penetrating showers can only explain about 1 per cent of the observed number of protons.

The most likely explanation of this excess of protons would seem to be that they are indeed produced by disintegrations, but that the protons have very much larger energy than those emitted from the stars observed in the emulsion. They must also be strongly collimated downwards, to explain the observed angular distribution. The protons must have a high energy to explain why few or none are observed to start in the emulsion. An estimate of the energy of production can be made as follows. Let L be the range in air of the protons at point of production, and let n protons be produced per cc./day. Suppose that all the protons are emitted downwards. Let d be the thickness of the emulsion and s the stopping power. Assume that the ratio of the rate of production in the emulsion rate of production in the air is also s . Also assume that all protons with a range of less than L_0 can be seen, and that all those with range greater than L_0 cannot.

Case 1. $L < L_0$. See fig. 7.

Fig. 7.



$$\text{Tracks crossing emulsion/cm.}^2/\text{day} = n(L - sd).$$

$$\text{Tracks stopping in the emulsion/cm.}^2/\text{day} = nsd.$$

$$\text{Tracks starting in the emulsion/cm.}^2/\text{day} = nsd.$$

$$\frac{\text{No. of tracks stopping/cm.}^2/\text{day}}{\text{No. of tracks crossing/cm.}^2/\text{day}} = \frac{sd}{L - sd}$$

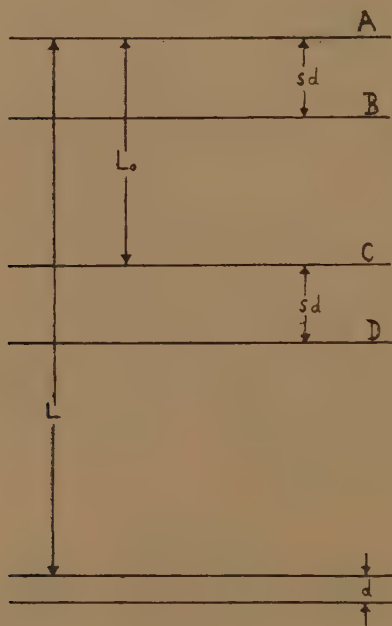
Case II. $L > L_0$.

$$\text{No. of tracks observed to cross emulsion/cm.}^2/\text{day} = n(L_0 - sd)$$

$$\text{No. of tracks stopping in the emulsion/cm.}^2/\text{day} = nsd.$$

$$\text{No. of tracks appearing to start in the emulsion/cm.}^2/\text{day} = nsd,$$

Fig. 8.



since tracks in AC will be observed to cross the emulsion, tracks in AB will stop in emulsion, and tracks in CD will appear to start in the emulsion (see fig. 8). In practice, the number appearing to start in the emulsion would be zero, since L_0 is not clearly defined.

This gives
$$\frac{\text{No. of tracks stopping/cm.}^2/\text{day}}{\text{No. of tracks observed to cross/cm.}^2/\text{day}} = \frac{sd}{L_0}$$

since $sd \ll L_0$.

Experimentally we find

$$\frac{\text{No. of tracks stopping/cm.}^2/\text{day}}{\text{No. of tracks observed to cross/cm.}^2/\text{day}} = 0.06.$$

Some of those stopping, however, are protons from the stars. Allowing for these, and also for those single protons observed to start in emulsion, we find

$$\frac{\text{No. of tracks stopping/cm.}^2/\text{day}}{\text{No. of tracks observed to cross/cm.}^2/\text{day}} = 0.025.$$

This gives a value of 4200μ for the range of the protons, corresponding to an energy of 32 MeV., at production. This is approximately equal to L_0 and therefore these results are consistent with the assumption that the protons are produced with energies sufficiently high for those starting in the emulsion to be unobserved.

The number of these fast protons produced/cc./day can be calculated as follows.

Suppose that all the protons are produced with a range $L > L_0$. Let $R(\theta) dv d\Omega$ be the number of protons emitted into a solid angle $d\Omega$, at angle θ with the vertical, from volume dv/day . Then

$$\begin{aligned} P(\theta) d\theta &= \int_{L-L_0}^L 2\pi l^2 \sin \theta d\theta \frac{\cos \theta}{l^2} \{R(\theta) + R(\theta + \pi)\} dl \\ &= 2\pi L_0 \sin \theta \cos \theta \{R(\theta) + R(\theta + \pi)\} d\theta. \end{aligned} \quad (9)$$

(see fig 1).

Now, from the definition of $R(\theta)$

$$\int_0^{\pi/2} 2\pi \sin \theta \{R(\theta) + R(\theta + \pi)\} d\theta = n, \quad (10)$$

where n is the total number of protons produced/cc./day. \therefore from (9) and (10)

$$\int_0^{\pi/2} \frac{P(\theta) d\theta}{L_0 \cos \theta} = \int_0^{\pi/2} 2\pi \sin \theta \{R(\theta) + R(\theta + \pi)\} d\theta = n \quad (11)$$

n can therefore be found from eqn. (11).

Taking L_0 as 4200μ in the emulsion, we find

$$n = 62 \text{ cc./day.}$$

Since 14 stars are observed/cc./day in the emulsion, there are 4.4 fast protons produced for every star. The minimum energy required to produce these is $4.4(32+8)=180$ MeV. (taking 8 MeV. as the binding energy). This is approximately equal to the average energy required to produce a star in the emulsion.

This can be checked by finding the range distribution of the protons observed in the plate. This should be constant up to a range of the order of L_0 , and then drop to zero. Measurements are being made to obtain the range distribution. Work is also being carried out to find the effect of changing the material immediately surrounding the plate, and the results will be published later.

ACKNOWLEDGMENTS.

I am deeply indebted to Professor G. P. Thomson for many valuable suggestions, and for his continued interest and advice.

REFERENCES.

- BERNARDINI, G., CORTINI, G., and MANFREDINI, A., 1948, *Physical Rev.*, **74**, October 1st.
BAGGE, E., *Cosmic Radiation* by Heisenberg, Chapter 13.
PERKINS, D. H., 1947 a, *Nature, Lond.*, **160**, 707; 1947 b, *Ibid.*, **160**, 299.
WIDHALM, A., 1940, *Z. Physik*, **115**, 481.

XXXVI. *The Primary Solid Solution of Silver in Aluminium.*

By Dr. G. V. RAYNOR and Dr. D. W. WAKEMAN,
Department of Metallurgy, University of Birmingham*.

[Received January 7th, 1949.]

SYNOPSIS.

The solid solubility curve for the primary solid solution of silver in aluminium has been re-determined, using pure materials. In contrast to the reports of previous workers, the solubility curve is not smooth, but exhibits a marked change in direction in the region of 49.4 per cent by weight of silver, at a temperature of 526° C. The maximum solid solubility is 55.6 per cent by weight of silver at the eutectic temperature of 566° C. The solidus curve shows an inflexion at approximately the same composition as that at which the anomaly in the solubility curve occurs.

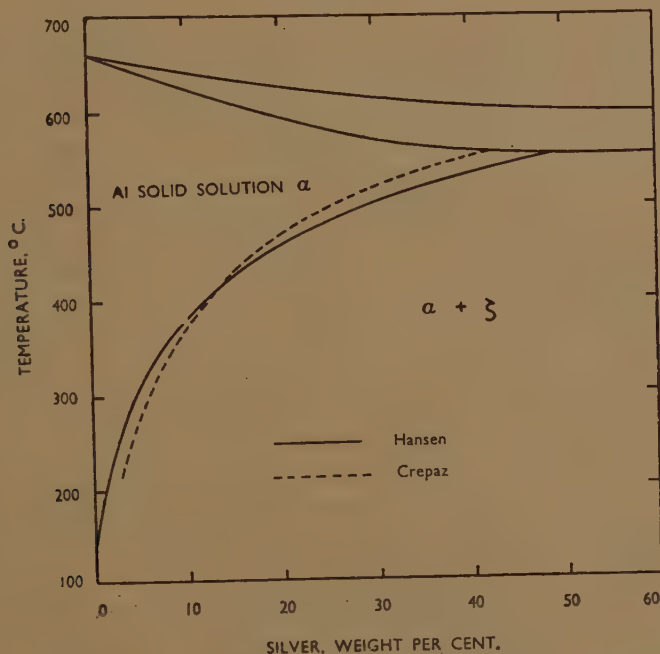
The results are discussed, and it is shown that the change in direction of the solubility curve occurs at the same electron : atom ratio as the peak of the ($\alpha + \alpha_1$) miscibility gap in the aluminium-zinc system. It is concluded that in both systems there is a tendency, caused by a factor of electronic origin, towards the separation of the solid solution into two face-centred cubic phases at an electron : atom ratio of 2.6. In the aluminium-silver system, however, the whole solubility curve lies at slightly too high a temperature for the formation of an ($\alpha + \alpha_1$) region such as occurs in the aluminium-zinc system, but the tendency towards this affects the form of the curve, introducing a change in direction. The relationship between the aluminium-zinc and aluminium-silver systems is described in terms of free-energy curves, and the structure of the Brillouin Zone for aluminium, which the present work suggests is still imperfectly known.

* Communicated by the Authors.

§ 1. INTRODUCTION.

THE general form of the equilibrium diagram for the aluminium-silver alloys has been satisfactorily established (Hansen 1936). There is an abnormally extensive solid solubility of silver in aluminium; the solubility decreases rapidly as the temperature falls. The primary solid solution, denoted α in this paper, enters into equilibrium with the ζ^* phase, the nature of which has been discussed by Hume-Rothery, Reynolds and Raynor (1940). The results reported by different workers (Tasaki 1927, Hansen 1928, Crepaz 1929, Guertler 1940), however, differ somewhat

Fig. 1.

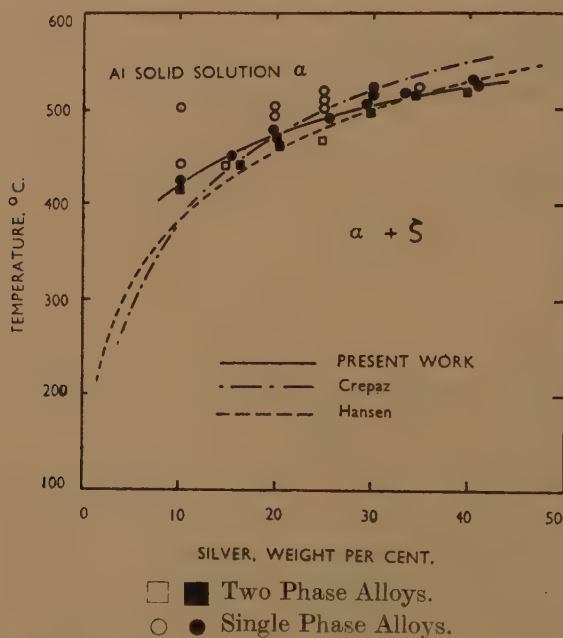


in detail, particularly with regard to the solid solubility limits, at various temperatures, in the aluminium-rich alloys. Fig. 1 shows the equilibrium diagram, up to 60 per cent of silver, as reported by Hansen (1928); the more recent work of Guertler (1940) is in substantial agreement with this. The broken curve represents the solubility values reported by Crepaz, and it may be noted that, according to Crepaz, the solubility is higher at low temperatures, and lower at high temperatures, than that given by Hansen. The aluminium used by Hansen was 99.74 per cent pure, and contained 0.1 per cent of silicon and 0.16 per cent of iron. The material employed by Crepaz was less pure, and contained only 99.3 per cent of aluminium. It therefore appeared to be advisable, in connection with a recent examination of the ternary aluminium-manganese-silver

* The nomenclature employed is that of Hume-Rothery, Reynolds and Raynor (1940).

system (Wakeman and Raynor 1948), to re-determine, in the appropriate temperature range, the solubility of pure silver in superpure aluminium. The results obtained, between 400° and 520° C., are summarized in fig. 2, in which the results of Hansen and Crepaz are included for comparison. The solubility of silver in aluminium is less at the lower temperatures, and higher at higher temperatures, than previously reported, and it is significant that the new curve bears a relationship to that of Hansen which is similar to the relationship of Hansen's curve to that of Crepaz. It is probable, therefore, that progressive improvement in purity of materials leads to a larger solid solubility of silver in aluminium at high

Fig. 2.



temperatures. Fig. 2 shows that, at the highest temperature investigated, the solid solubility curve becomes almost horizontal; extrapolation of the curve to the temperature of the liquid $\rightleftharpoons \alpha + \zeta$ eutectic would indicate an impossibly high maximum solid solubility. Further experiments on the system were therefore made, and are reported in the present paper.

§ 2. MATERIALS USED, AND EXPERIMENTAL METHODS.

Alloys containing from 10 to 60 per cent by weight of silver were prepared from assay silver and superpure aluminium. The silver was obtained from Messrs. Johnson Matthey & Co. Ltd., and was 99.99 per cent pure, while the aluminium was supplied by the British Aluminium Co. Ltd. The metals were melted in alumina-lined crucibles, well stirred with alumina rods, and cast into a chill mould. Annealing treatments (3 to 10 days in duration) were carried out *in vacuo* in controlled resistance furnaces, and were concluded by quenching into cold water. In general,

specimens were obtained in a homogeneous condition and subsequently re-annealed at lower temperatures to secure precipitation of the ζ phase. All critical microspecimens were chemically analysed.

The solidus experiments were carried out on alloys which had been homogenized by annealing at 548° C. for 70 hours. Specimens were held at higher temperatures for periods varying from $2\frac{1}{2}$ to $6\frac{1}{2}$ hours and quenched in cold water.

The micrographic examination of the quenched alloys followed conventional lines, and the recognition of small traces of the ζ phase, even in the unetched specimens, was not difficult. The presence of chilled liquid was also relatively easy to establish.

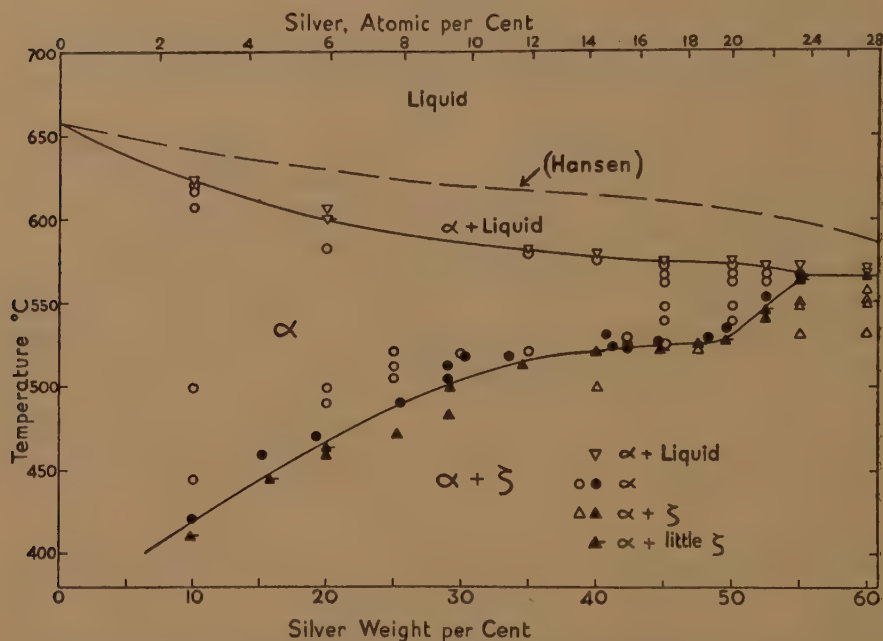
Polishing was carried out in the manner described previously (Wakeman and Raynor 1948), the final polish being applied with "Silvo" metal polish on Selvyt cloth. The ζ phase could be distinguished in the as-polished microsections as faint grey particles; better distinctions were obtained by exposing the specimens to the atmosphere for 12 to 24 hours, when the ζ phase darkened considerably. Chilled liquid, in alloys quenched from above the solidus, was best revealed by etching for 10 seconds in an aqueous solution containing 0.5 per cent of hydrofluoric acid, washing, and cleaning in a solution containing the chromate ion. The composition of the latter solution was not critical; the composition most generally used was a saturated solution of potassium bichromate in 10 per cent sulphuric acid.

§ 3. EXPERIMENTAL RESULTS.

The results obtained are shown in fig. 3; analysed alloys are distinguished from unanalysed specimens, which have been denoted by open symbols. The slowly rising portion of the curve shown in fig. 2 is confirmed by the new results, and it may be further noted that a marked change in direction occurs at 49.4 per cent by weight of silver, and in the region of 526° C. This feature has not previously been observed, owing, probably, to the inferior purity of the materials used. The system has been studied in detail over the composition range close to the change of direction of the solubility curve, and the results admit of no other interpretation. The curve cannot be a smooth one as reported by Hansen and by Crepaz. The solidus curve at first falls smoothly as silver is added to aluminium, but, in the region 30 to 40 per cent by weight of silver, begins to flatten out, and becomes almost horizontal within the same composition range as that occupied by the almost horizontal portion of the solubility curve. The solidus curve then falls smoothly to cut the eutectic horizontal. According to the micrographic experiments, the eutectic temperature lies between 566° C. and 568.5° C.; at the latter temperature, much chilled liquid was visible in the microstructure. A cooling curve, using an alloy containing 67 per cent by weight of silver, gave a strong eutectic arrest at 565.7° C., in good agreement with the micrographic work. This value is also in good agreement with the temperatures of 567° and 568° C., reported respectively by Petrenko (1905) and Crepaz (1929), but does not confirm the value of 558° C. given by Hansen (1928).

The maximum solid solubility of silver in aluminium is 55.6 per cent by weight, as compared with 48 per cent according to Hansen (1928). It may be noted (fig. 3) that the alloy containing 55.0 weight per cent of silver (by analysis) was obtained as a homogeneous specimen at 567° C., but contained the ζ phase at 562.3° C. The point of maximum solubility

Fig. 3.



is therefore very accurately located, and could hardly lie outside the extreme limits 55.3 to 55.8 weight per cent of silver.

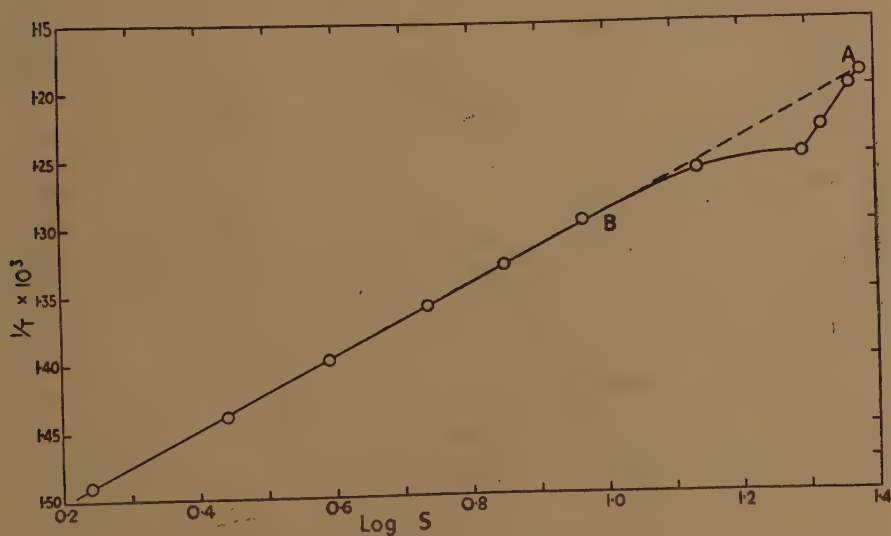
§ 4. DISCUSSION.

The present work makes it clear that the solid solubility of pure silver in high purity aluminium does not decrease smoothly from the maximum value as the temperature falls. Instead, there is a marked change in direction in the region of 49.4 per cent by weight of silver and 526° C. At approximately the same composition, the solidus curve becomes almost horizontal. According to the micrographic results in the solid state, the solubility curve below 526° C. never becomes quite horizontal, but maintains a small positive slope up to the point at which the subsequent rapid rise occurs. During the micrographic work, no difference was observed in the appearance, etching behaviour or mode of precipitation of the ζ phase above and below the critical temperature of 526° C., and duplex microstructures obtained from these two temperature ranges proved indistinguishable. Similarly, no differences in the characteristics of the α phase were observed above and below the critical temperature. There is no

suggestion in the literature that the ζ phase itself undergoes any transformation in the region of 526°C ., and the aluminium-rich boundary of this phase shows no anomaly at this temperature (Hume-Rothery, Reynolds and Raynor 1940). The anomalous character of the primary solid solubility boundary must therefore be a characteristic of the solid solution itself.

In fig. 4, the solubility data have been plotted as $\log S$ against $1/T$, where S is the solubility, in atomic per cent, at the temperature $T^\circ\text{A}$. Up to a $\log S$ value of 1.0 ($S=10$ atomic per cent), a satisfactory straight line is obtained, so that it is improbable that any lack of equilibrium affected the results in this composition range. Above $\log S=1.0$, the

Fig. 4.

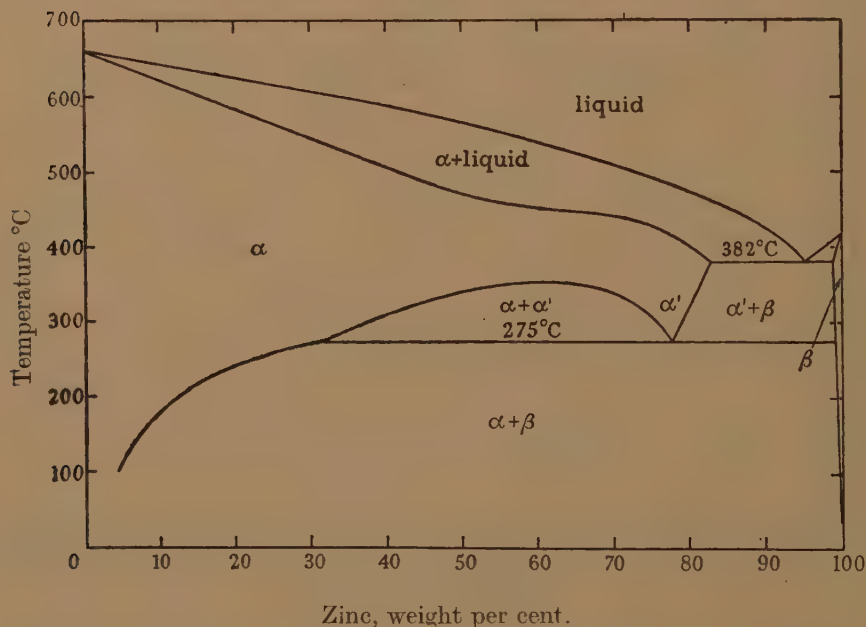


curve decreases gradually in slope, becoming almost horizontal, until the high temperature, steeply sloping portion of the boundary is reached. It may be noted that the extrapolation of the straight line given at low temperatures passes exactly through the point A, which represents the maximum solubility of silver in aluminium at the eutectic temperature of 566°C . Since equilibrium was established at the lower temperatures, it is improbable that the results at high temperatures, for which comparable times of annealing were used, are due to any lack of equilibrium. The deviation from linearity in the region BA of fig. 4 appears, therefore, to be due to some factor affecting the actual stability of the primary solid solution itself, and the difference between these results and those of previous workers can only be explained in terms of the purity of the materials used, to which the details of the equilibrium diagram appear to be very sensitive. The form of the curve in fig. 4 is not typical of systems in which the change in the direction of the solubility curve is

due to some transformation in the phase with which the primary solid solution is in equilibrium. In such cases, the change in direction is manifested on curves of $\log S$ against $1/T$ as a sharp intersection between two straight lines. As an example of this, the system magnesium-indium may be quoted (Hume-Rothery and Raynor 1938).

The general appearance of the aluminium-rich equilibrium diagram shown in fig. 3 is reminiscent of that of the aluminium-zinc equilibrium diagram, particularly as regards the flattened portion of the solidus curve. The aluminium-zinc diagram is reproduced for convenience as fig. 5.

Fig. 5.



Detailed comparison of the relevant portions of the two equilibrium diagrams shows that the change in direction of the aluminium-silver solubility curve, and the peak of the $(\alpha + \alpha_1)$ miscibility gap in the aluminium-zinc alloys occur at the same ratio of valency electrons to atoms. This may be seen from fig. 6, where the two solid solubility curves are plotted in terms of the electron : atom ratio, and which strongly suggests that the same factor is responsible for the somewhat anomalous form of the solubility relations in both systems. This factor must also have its origin in the electronic structure of the alloys. The correlation demonstrated in fig. 6 implies also that the almost horizontal portions of the two solidus curves occur at approximately the same electron : atom ratio.

It is of interest to note that the general form of diagram shown by the aluminium-silver alloys may be derived diagrammatically from that of the aluminium-zinc alloys, as shown in fig. 7. The existence of an

Fig. 6.

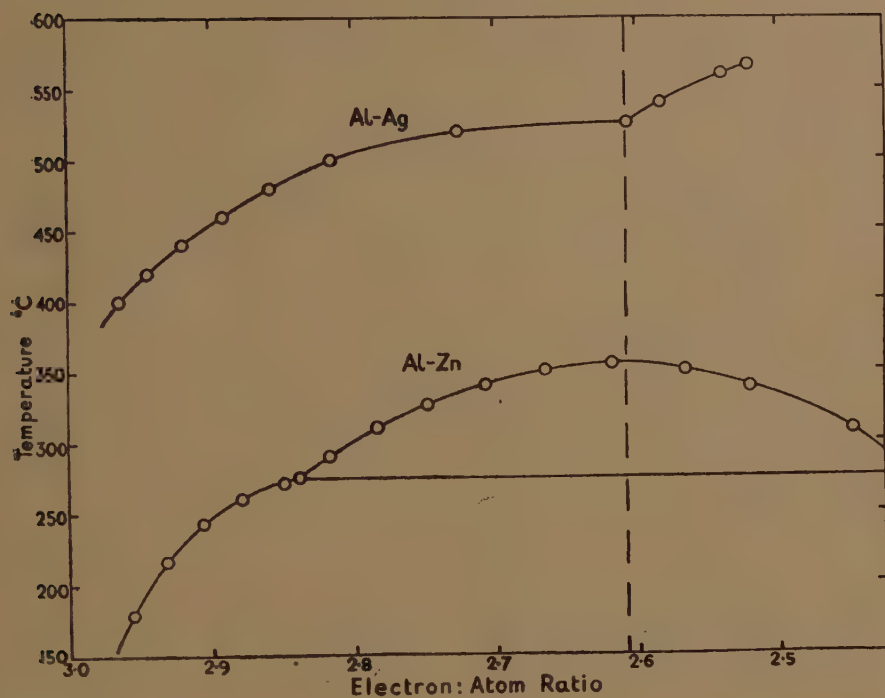
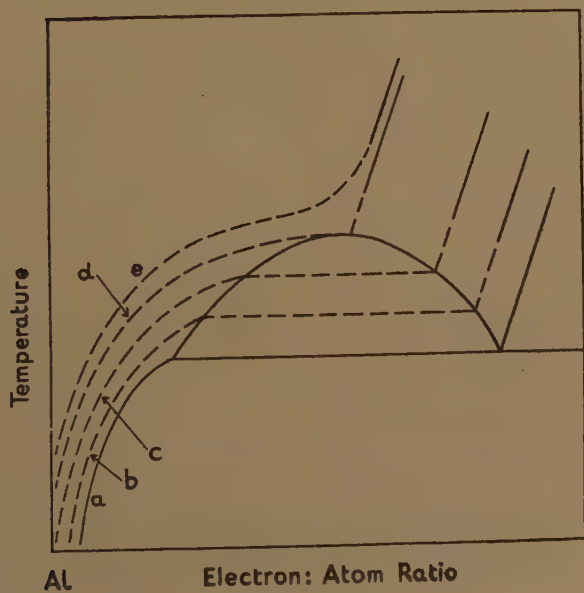


Fig. 7.



$(\alpha + \alpha_1)$ region of the form shown indicates that, as the temperature rises, the tendency towards immiscibility decreases until, at a certain critical temperature, complete miscibility is shown. If, now, the low temperature portion of the curve is considered to move towards higher temperatures (curve *b*), then the $(\alpha + \alpha_1)$ region becomes more limited in extent, and a smaller composition difference separates the end of the low temperature portion and the beginning of the high temperature portion of the solubility curve. Further rises in the temperature of the initial portion of the solubility curve produce curve *c*, and, in the limit, curve *d*. Further increase beyond this would be expected to give rise to a solubility curve such as curve *e*, containing a marked inflexion. The form of the aluminium-silver solubility curve is very close to that of curve *d* in fig. 7, from which it may be inferred that the critical temperature for the formation of an $(\alpha + \alpha_1)$ type of miscibility gap in aluminium-silver alloys is only slightly below 526°C. , at which the change in direction occurs. According to this interpretation, the anomalous form of the aluminium-silver equilibrium diagram is due to a tendency towards the formation of an $(\alpha + \alpha_1)$ region, which is countered by the relatively high temperature at which the solubility curve lies. The whole curve lies at slightly too high a temperature for the actual manifestation of the $(\alpha + \alpha_1)$ region in this system, but the tendency towards this is clear.

The formation of the $(\alpha + \alpha_1)$ region of the equilibrium diagram of the aluminium-zinc alloys has been discussed by Hume-Rothery (1944, 1946), in terms of hypothetical free-energy curves. Thus, the form of the free-energy curves for aluminium-zinc alloys at a temperature between 275°C. and the peak of the $(\alpha + \alpha_1)$ miscibility gap would be as shown in fig. 8 *a*, the irregularity in the curve for the primary solid solution of zinc in aluminium being responsible for the $(\alpha + \alpha_1)$ equilibrium. At higher temperatures, the irregularity is smoothed out (fig. 8 *b*) and a normal type of equilibrium is obtained. For the aluminium-silver alloys a somewhat modified free energy diagram is necessary, since the free-energy trough associated with the ζ phase will be deeper, relative to the aluminium-rich curve, than that associated with the solid solution of aluminium in zinc. This is reflected by the fact that the heats of mixing in the aluminium-zinc system are negative, whereas those in the aluminium-silver system rise to a maximum of $+1.1 \text{ k.cal./g. atom}$ at approximately the composition of the ζ phase. The appropriate form of free energy diagram for the aluminium-silver alloys at the lower temperatures is shown in fig. 9 *a*; the tendency towards the formation of an $(\alpha + \alpha_1)$ region is represented by the irregularity on the free energy curve for the primary solid solution, but, owing to the depth of the trough in the ζ curve, equilibrium involves a solid solution composition which is too small for the $(\alpha + \alpha_1)$ separation to be manifested. As the temperature rises, we may consider the irregularity to be smoothed out somewhat, and the condition of the alloys when the re-entrant portion has just been removed may be represented by fig. 9 *b*. Owing to the relative changes of the minima on the free energy scale with

Fig. 8.

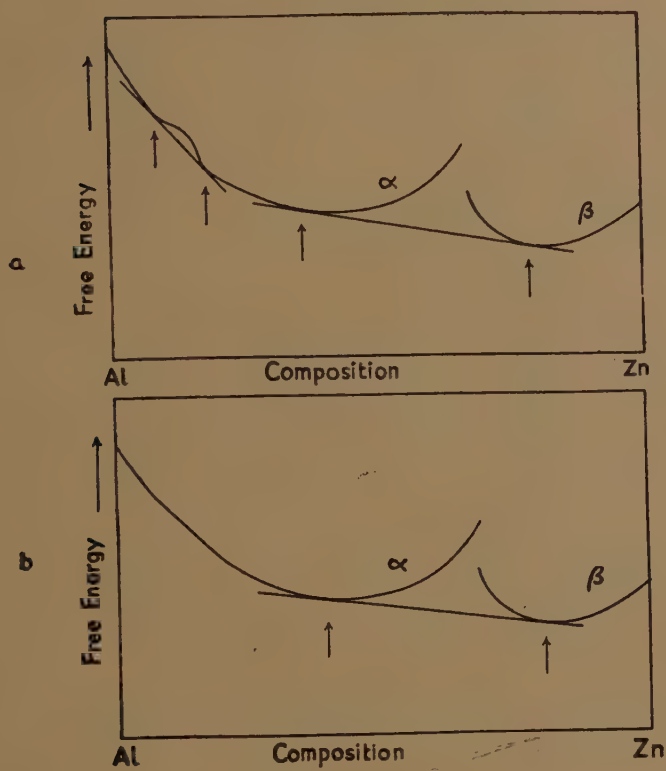


Fig. 9 a:

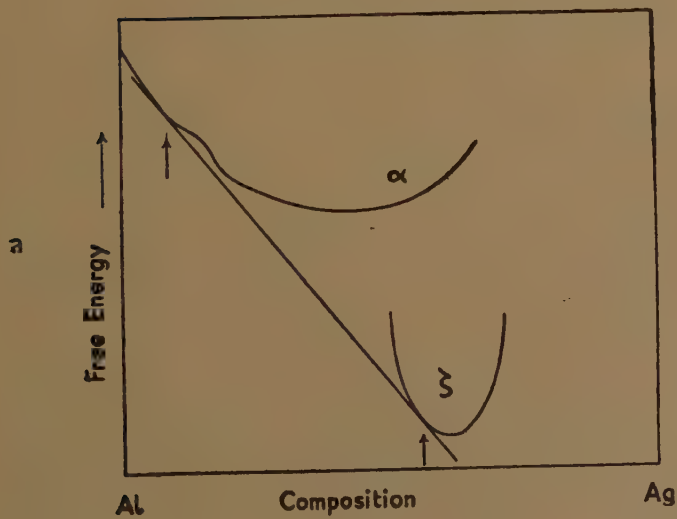
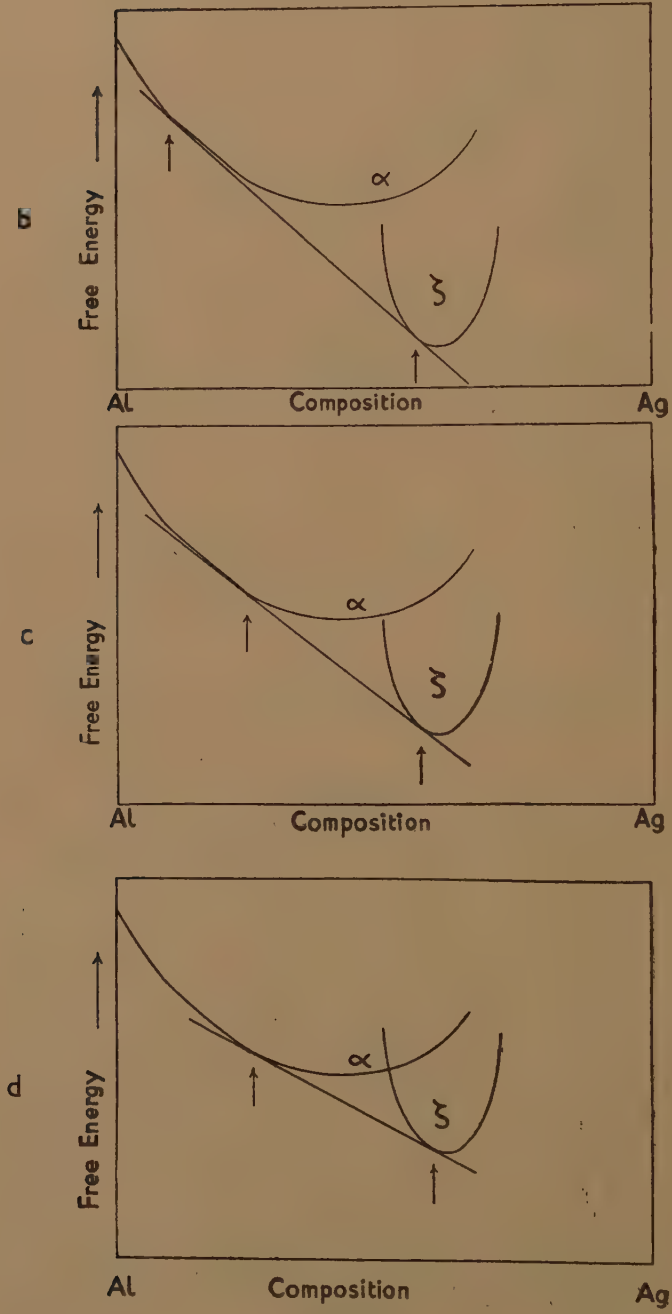


Fig. 9 b-d.



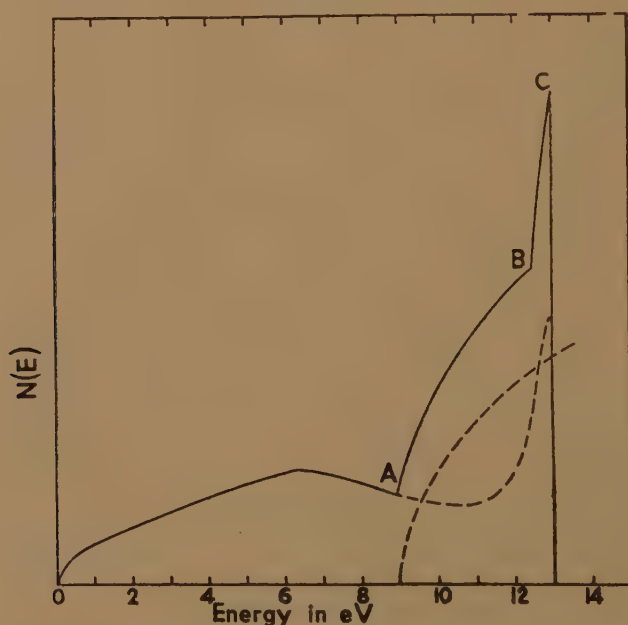
change in temperature, the solid solubility has increased. At this stage, a very small further change in the temperature, and hence in the relative positions of the free energy maxima, will cause a large change in the primary solid solubility, as indicated in fig. 9 *c*. The smoothing out of the irregularity in the free energy curve with rise in temperature thus allows an interpretation of the form of the solubility curve of silver in aluminium below the temperature of 526°C . At still higher temperatures, a normal type of equilibrium may be expected, as in fig. 9 *d*, and further temperature increases will not abnormally increase the solubility. It will be observed that the presence of the irregularity on the free energy curve has no influence on low temperature equilibria represented by fig. 9 *a*, and very little influence on the high temperature equilibria represented by fig. 9 *d*. The fact that the extrapolation of the low-temperature straight line of fig. 4 touches the point corresponding to the maximum solid solubility of silver in aluminium is consistent with this.

From the above discussion, it is clear that the free energies of the primary solid solutions of both zinc and silver in aluminium are modified by the introduction of an irregularity in the free-energy/composition curve, caused by a factor of electronic origin, at 2.6 electrons per atom. The forms of the solid solubility curves may be satisfactorily interpreted in terms of the presence of this irregularity. The nature of the factor which introduces the irregularity is not, however, clear. On the basis of $N(E)/E$ curves for the face-centred cubic structure of copper, published by Jones and Mott (1937), it has been suggested (Hume-Rothery 1944, 1946), that the electron : atom ratio at which the irregularity in the free energy curve for the aluminium-zinc primary solid solution occurs corresponds with a relatively deep trough in the $N(E)$ curve. Since this suggestion was made, however, an examination of the electronic structure of aluminium has been made by Matyáš (1948), and the new $N(E)$ curve, taken from the work of Matyáš, has been described by Axon and Hume-Rothery (1948). The form of this curve, which is reproduced here as fig. 10, is strikingly different from that given for copper by Jones and Mott, and indicates that the latter curve cannot be used in connection with aluminium alloys. The curve of fig. 10 shows a small trough A at an electron : atom ratio which is much too small to be significant in the present work. The actual electron : atom ratio at which the irregularity in the free energy curves arises (2.6 electrons per atom) corresponds approximately with the point B of fig. 10, that is, the irregularity occurs when the area under the sharply rising peak BC has been emptied of electrons. It is most significant that the electron : atom ratio at which the change in direction of the solubility curve for silver in aluminium occurs should correspond with a break in the $N(E)$ curve for aluminium, but it is somewhat difficult to reconcile the form of the $N(E)$ curve with the required form of the free-energy curves. It is possible that the $N(E)$ curve for aluminium reproduced in fig. 10, though qualitatively correct, may still be incorrect in detail. If, for example, the broken curve for the 1st zone above 9 e.V. were to fall to a

sharp minimum in the region of 12 to 12.5 e.V., instead of to a shallow minimum at 11 e.V., a trough might be introduced into the summation curve at a point corresponding roughly to 2.6 electrons per atom.

The general conclusion from this comparison of the forms of the aluminium-silver and aluminium-zinc equilibrium diagrams is that the anomalies in both the primary solid solubility curves are due to the same factor, which is of an electronic origin. At the same electron : atom ratio (2.6) in each case, there must be an irregularity in the free-energy curves corresponding to the primary solid solution. Consideration of the energy distribution of electrons in solid aluminium indicates that the electron :

Fig. 10.



atom ratio of 2.6 corresponds approximately with a kink in the $N(E)$ curve ; it is considered, however, that the details of the $N(E)$ curve are not yet known with sufficient certainty for the correlation to proceed any further.

ACKNOWLEDGMENTS.

This research was carried out under the general supervision of Professor D. Hanson, D.Sc., to whom the authors' thanks are due for his interest and support. The authors must also acknowledge the valuable assistance of Mr. G. Welsh, and of Mr. A. J. Hawkes, who carried out the chemical analyses. Grateful acknowledgment is also made to the Department of Scientific and Industrial Research, the Royal Society, the Chemical Society, and Imperial Chemical Industries, Ltd., for generous financial assistance.

REFERENCES.

- AXON, H. J., and HUME-ROTHERY, W., 1948, *Proc. Roy. Soc. A*, **193**, 1.
 CREPAZ, E., 1929, *Atti. III Congr. naz. Chim. pura appl. Firenze e Toscana*, 371.
 GUERTLER, W., 1940, *Metallwirtschaft*, **19**, 432.
 HANSEN, M., 1928, *Z. Metallkunde*, **20**, 217; 1936, *Der Aufbau der Zweistofflegierungen* (Berlin: Julius Springer).
 HUME-ROTHERY, W., and RAYNOR, G. V., 1938, *J. Inst. Metals*, **63**, 227.
 HUME-ROTHERY, W., REYNOLDS, P. W., and RAYNOR, G. V., 1940, *J. Inst. Metals*, **66**, 191.
 HUME-ROTHERY, W., 1944, *J. Inst. Metals*, **70**, 229.; 1946, *Inst. Metals Monograph and Report Series*, No. 2.
 JONES, H., and MOTT, N. F., 1937, *Proc. Roy. Soc. A*, **162**, 49.
 MATYÁŠ, R., 1948, *Phil. Mag.* [7], **39**, 429.
 PETRENKO, G. J., 1905, *Z. Anorg. Chem.*, **46**, 49.
 TASAKI, M., 1927, *Kinzoku no Kenkyu*, **4**, 35.
 WAKEMAN, D. W., and RAYNOR, G. V., 1948, *J. Inst. Metals*, **75**, 131.

XXXVII. Channel Section Waveguide Radiator.

By A. L. CULLEN, B.Sc.,

Electrical Engineering Department, University College, London*.

[Received October 6, 1948.]

SUMMARY.

The transmission characteristics of a channel section waveguide radiator are studied theoretically, and formulæ for its phase and attenuation coefficients are derived. The theory is supported by experimental results.

§ 1. INTRODUCTION.

THE channel section waveguide whose characteristics are studied in this paper is shown in fig. 1(a). In fig. 1(b) is shown a rectangular section waveguide supporting the H_{01} -mode. Provided that $a \ll \lambda$, it is reasonable to assume that the power lost by radiation from the open side of the channel section will not be very great, and to a first approximation, we might expect to find a mode of propagation in which the field within the channel is similar to that in half of the wave guide of fig. 1(b). We may, in fact, consider the rectangular guide of fig. 1(b) as built up from two channel sections with their open sides fitted together in such a way that one channel forms a fringing guard for the other. For such a guide we have

$$h = k\sqrt{1 - (\lambda/2b)^2}, \quad \dots \dots \dots (1)$$

in which $h = 2\pi/\lambda_g$ and $k = 2\pi/\lambda$. In the absence of such a guard, fringing

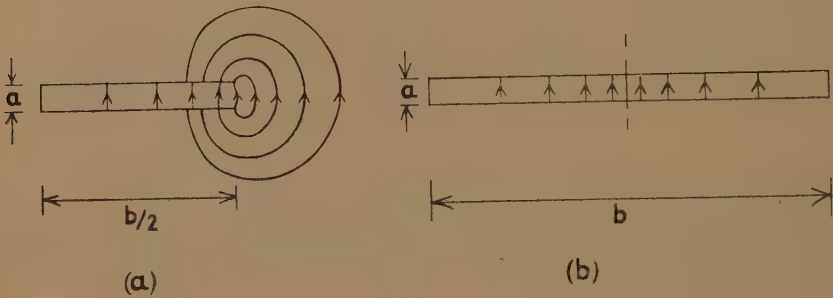
* Communicated by the Author

will occur, as indicated in fig. 1(a). It is important to note that the fringing field is electromagnetic and not purely electric, as in the well known case of the parallel plate condenser, and it is the presence of the external magnetic field which is responsible for the leakage of power. However, in the immediate vicinity of the open face the field energy is predominantly electric, and we shall find that the electric energy storage, which represents fringing capacitance can be calculated with sufficient accuracy by electrostatic theory. The effect of this additional stored electric energy on the phase constant may be allowed for by adding to the channel depth $b/2$ an edge correction $\delta/2$, whose value we shall determine, and calculating the phase constant from the usual formula for a rectangular guide of width b'

$$h = k\sqrt{1 - (\lambda/2b')^2}, \quad (2)$$

in which $b' = b + \delta$.

Fig. 1.



Channel section waveguide and equivalent rectangular guide.

Next, let us consider the mechanism of radiation at the open side, to get a closer approximation to actual propagation conditions.

Referring to fig. 2, consider two elements of the open side at P_1 and P_2 , separated by one guide wavelength. These two elements are always in phase, and so their contributions to the wave front passing through P_1 and distant one wavelength from P_1 will also be in phase.

Another way of looking at this effect is to split the wave in the guide into the usual two zig-zag components, and it is then clear that if that component plane wave which is travelling towards the open side is imperfectly reflected, the transmitted wave will emerge at the angle of incidence $\theta = \sin^{-1}(\lambda/\lambda_g)$ which leads to the previous result.

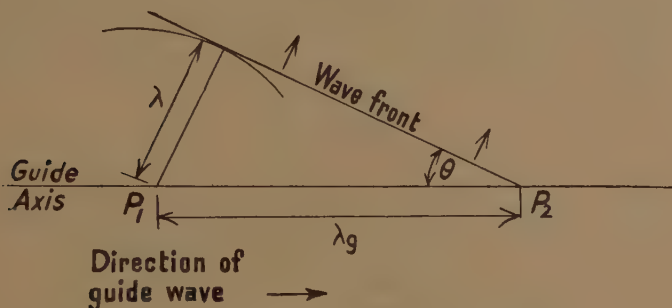
The emergent wave will not, of course, be plane, but will spread out like the fringing field to form a conical outward-travelling wave. The energy carried away from the guide by this process represents attenuation of the power flow within the channel, and may be allowed for by an attenuation coefficient α_r due to radiation.

If metal losses are significant, they may be allowed for by adding to α_r a coefficient α_m , so that the total attenuation coefficient α , is given by

$$\alpha = \alpha_r + \alpha_m. \quad \dots \dots \dots (3)$$

The evaluation of α_r and α_m will be considered later.

Fig. 2.



Formation of conical wave.

$$\text{If } h = 2\pi/\lambda g \quad \text{and} \quad k = 2\pi/\lambda,$$

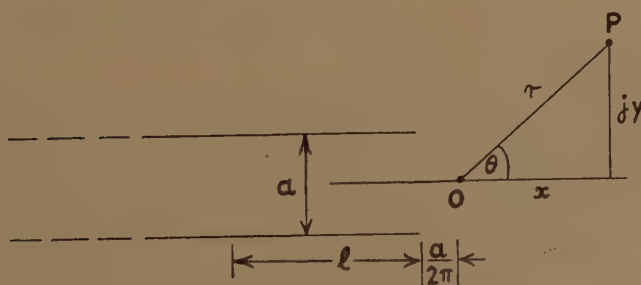
$$\text{Then } \sin \theta = \lambda/\lambda g.$$

$$\cos \theta = \sqrt{1 - \left(\frac{\lambda}{\lambda g}\right)^2} = \sqrt{1 - \frac{h^2}{k^2}}.$$

§ 2. EFFECT OF FRINGING ON PHASE CONSTANT.

In fig. 3, two semi-infinite plates separated by a distance " a ", and having a difference in potential of V , are arranged to form a parallel plate condenser. We shall study the fringing field of such a condenser

Fig. 3.



Coordinate system in z -plane.

with the object of applying the results to the wave guide problem discussed in the previous section. We take an origin on the zero-equipotential plane, distant $a/2\pi$ from the open face of the guide, and use complex coordinates to specify the point P at $z = x + jy$. Then it can be

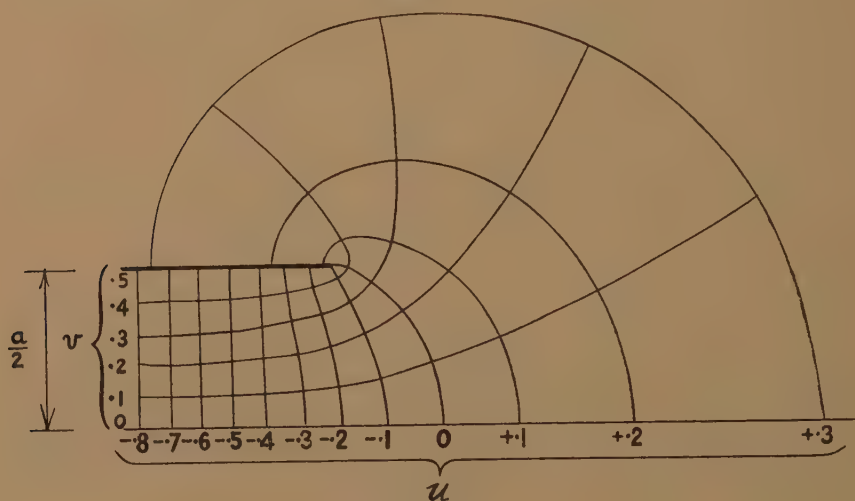
shown that the equipotential lines and lines of force are given by putting $v=\text{constant}$, and $u=\text{constant}$ respectively in the equation

$$z = \frac{a}{2\pi} \left[\exp \left(\frac{2\pi}{V} [u + jv] \right) + \frac{2\pi}{V} [u + jv] \right]. \quad (4)$$

The result is shown in fig. 4. Note that u , which corresponds to the electric flux, is positive outside the plates except in the immediate vicinity of the open side, and negative inside. It is clear from (4) that for large negative values of u the exponential term becomes extremely small and we have the simple relationship

$$z \doteq \frac{a}{V} [u + jv]. \quad (4a)$$

Fig. 4.



Lines of force and equipotentials in fringing field.

This gives an orthogonal set of straight lines for the lines of force and equipotential surfaces, which corresponds with physically obvious fact that the field well inside the semi-infinite condenser is approximately the same as that in an infinite condenser. For large positive values of u we have the approximate relationship

$$z \doteq \frac{a}{2\pi} \exp \left(\frac{2\pi}{V} [u + jv] \right), \quad (4b)$$

and this is valid at great distances from the open side. From (4b) we deduce that the flux function u at a great distance $r=|z|$, ($|z| \gg a$) from the origin is given by

$$u = \frac{V}{2\pi} \log_e \left(\frac{2\pi r}{a} \right). \quad (5)$$

The tangential component of the electric field may be calculated from (5) if we remember that κu = linear electric flux density (where κ is the absolute permittivity of the medium) so that the surface electric flux density D_φ is given by $\kappa(\partial u/\partial r)$. But $E_\varphi = D_\varphi/\kappa$ whence

$$E_\varphi = \frac{\partial u}{\partial r}. \quad \dots \dots \dots (6)$$

Combining (5) and (6), we find the tangential electric field at great distances is given approximately by

$$E_\varphi = \frac{V}{2\pi r}, \quad \dots \dots \dots (7)$$

having an obvious geometrical interpretation, the lines of force of E being approximately circular of radius r . It is important to note that the criterion of validity for equation (7) is $r \gg a$.

It is not difficult to show that the effect of fringing is to *increase* the charge stored on the inner surfaces of the plates, and also to store additional charge on the outer surfaces. The increased internal charge is electrically equivalent to adding strips of width $a/2\pi$ to top and bottom plates. The external charge, however, is infinite for infinite plates, so that the fringing capacitance of such plates is infinite. However, if we consider a finite width of plate, say l (see fig. 3), it can be shown that the charge storage is equivalent to a further added strip of width $(a/2\pi \log_e (2\pi l/a))$.

Now returning to our channel radiator, it is clear that conditions are somewhat different. In the first place, the field is alternating rapidly. Indeed, if this were not the case, no P.D. could exist across the open face. In the second place the fact that the outer surfaces are not infinite, and share one edge in common, electrically, will considerably modify the charge distribution. Nevertheless, we shall assume that a reasonably good approximation to the value of $\delta/2$, the "fringing" strip, will result if we put $l=b/2$ in the preceding results.

This leads to

$$\delta = \frac{a}{\pi} \left[1 + \log_e \frac{\pi b}{a} \right]. \quad \dots \dots \dots (8)$$

Remembering that the effective breadth b' of the equivalent guide is $b + \delta$, and using equation (2), we are now in a position to calculate approximately the phase constant of the channel section waveguide. Experimental results confirming (8) will be described in § 5.

§ 3. CALCULATION OF RADIATED POWER.

Let us now turn to the problem of calculating the power radiated from the open face of the guide. This may be accomplished by the usual Poynting vector method.

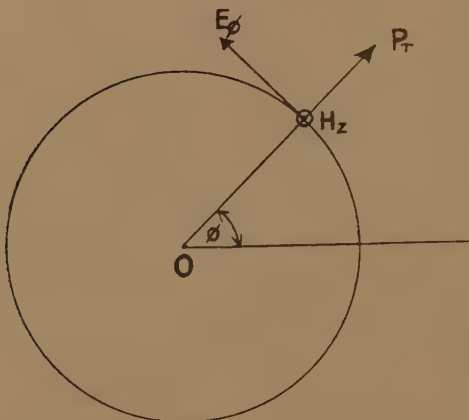
The power radiated away from the guide per unit length is equal to the real part of the radial component of the complex Poynting vector integrated over a unit length of any cylindrical surface surrounding the guide (fig. 5). Thus if dP_r is the power radiated from a length dz of the guide, we have

$$\left. \begin{aligned} dP_r &= \text{Re} \, dz \int_0^{2\pi} E_\phi H_z^* r \, d\phi \\ \text{or} \quad \frac{dP_r}{dz} &= \text{Re} \int_0^{2\pi} E_\phi H_z^* r \, d\phi. \end{aligned} \right\} \dots \dots \dots (9)$$

It will be convenient, as usual, to take a large value of r for this simplifies the analysis. Now it can be shown that the most general cylindrical wave can be derived from a set of scalar wave functions ψ_n , given by

$$\psi_n = e^{-jn\phi} H_n^{(2)}(\sqrt{k^2 - h^2} \cdot r) e^{-jh z} \dots \dots \dots (10)$$

Fig. 5.



Radial Poynting vector used in calculation of power radiated.

In this expression n has integral values only, and $H_n^{(2)}$ is the Hankel function of the second kind, which, with a time factor $e^{j\omega t}$ represents an outward-travelling cylindrical wave. E_ϕ and H_z may be expanded in series of such function thus

$$\left. \begin{aligned} E_\phi &= \sum_{n=-\infty}^{+\infty} A_n \frac{\partial \psi_n}{\partial r}, \\ H_z &= \frac{k^2 - h^2}{j\omega\mu} \sum_{n=-\infty}^{+\infty} A_n \psi_n. \end{aligned} \right\} \dots \dots \dots (11)$$

If we imagine the field outside the channel section waveguide to be expanded in this way it is clear that the resulting formulæ must reduce

to the electrostatic fringing field under appropriate conditions. In the first place, the electrostatic field is two dimensional, having no variation with z , so we must put $h=0$ in (10). If we restrict our attention to values of r such that the condition $r \gg a$ is fulfilled, the electrostatic fringing field is approximately circularly symmetrical, so we may neglect the higher order terms in the summations in (11), retaining only the principal terms in each expansion thus, for the electric field

$$\left. \begin{aligned} E_{\varphi} &= A_0 \frac{\partial \psi_0}{\partial r}, \\ \psi_0 &= H_0^{(2)}(kr). \end{aligned} \right\} \dots \dots \dots (12)$$

where

We now introduce the restriction, necessary for the quasi-static case, that $\lambda \gg r$. (The theory given is then valid only if a value of r can be found such that $\lambda \gg r \gg a$, a restriction which, though difficult to meet accurately, is satisfied well enough to give useful approximate formulæ in many cases, as we shall see). This restriction makes $kr \ll 1$, and the zero order Hankel function may be simplified to give

$$\psi_0 = 1 - j \frac{2}{\pi} [\log_e kr + C - \log_e 2], \dots \dots \dots (13)$$

in which C is Euler's constant, approximately 0.577. Using this value of ψ_0 in (12), we find

$$E_{\varphi} = -j A_0 \frac{2}{\pi r} \dots \dots \dots (14)$$

Comparing this value of E_{φ} with that given by eqn. (7), we see that A_0 may be expressed in terms of the voltage V across the open side of the guide.

$$\left. \begin{aligned} -j A_0 &= \frac{V}{4} \\ A_0 &= +j \frac{V}{4}. \end{aligned} \right\} \dots \dots \dots (15)$$

We may use this value of A_0 for all frequencies for which a value of r exists such that $\lambda \gg r \gg a$. It is important to note that this is not a restriction on r , and values of $r \sim \lambda$ may be dealt with using the value of A_0 given by (15) in equation (12). In particular, the field at $r \gg \lambda$ may be found by using the large-argument approximation to the Hankel function, *i. e.*

$$\psi_0 \doteq \sqrt{\frac{2}{\pi k_1 r}} e^{-j(k_1 r - \frac{\pi}{4})}, \dots \dots \dots (16)$$

where $k_1 = \sqrt{k^2 - h^2}$.

Neglecting the induction field, E_ϕ and H_z are found to be

$$\left. \begin{aligned} E_\phi &= \frac{k_1 V}{4} \sqrt{\frac{2}{\pi k_1 r}} \cdot e^{-j(k_1 r - \frac{\pi}{4})}, \\ H_z &= \left(\frac{k^2 - h^2}{k} \right) \frac{V}{4} \sqrt{\frac{\kappa}{\mu}} \sqrt{\frac{2}{\pi k_1 r}} \cdot e^{-j(k_1 r - \frac{\pi}{4})}. \end{aligned} \right\} \dots \dots (17)$$

It follows from (17) that

$$E_\phi H_z^* = \frac{V^2}{8\pi} \sqrt{\frac{\kappa}{\mu}} \left(\frac{k^2 - h^2}{kr} \right) \dots \dots \dots (18)$$

Substituting this result in equation (9), noticing that it is independent of ϕ , we find

$$\frac{dP}{dz} = \frac{V^2}{4} \sqrt{\frac{\kappa}{\mu}} \left(\frac{k^2 - h^2}{k} \right), \dots \dots \dots (19)$$

which is the required expression for the power radiated per unit length of guide. This result will be applied to the calculation of the effective attenuation coefficient α_r due to radiation.

§ 4. CALCULATION OF ATTENUATION COEFFICIENT.

We shall calculate the attenuation coefficient by making use of the general formula

$$\alpha = \frac{1}{2P} \cdot \frac{dP}{dz}, \dots \dots \dots (20)$$

in which P is the power flow past a cross section at z and dP/dz is the power loss per unit length. This factor will contain two components, power lost by radiation, and power lost in the walls of the channel. Thus, the total attenuation coefficient can be written as the sum of α_r , due to radiation, and α_m , due to wall losses.

Noticing that the channel section has half the wall loss and half the power flow of a complete rectangular guide, for a given field intensity, we see that α_m for the channel is given by the usual formula for a rectangular guide, namely

$$\alpha_m = \frac{\pi \Delta \lambda_g}{\lambda^2} \left[\frac{1}{a} + \frac{\lambda^2}{2b^3} \right], \dots \dots \dots (21)$$

in which Δ is the penetration depth in the walls of the guide.

To calculate α_r , note that the power flow down the guide is

$$P = \frac{1}{2} \frac{V^2}{Z_{w,v}}, \dots \dots \dots (22)$$

where $Z_{w,v}$ is the power-voltage characteristic impedance of the complete rectangular guide, which is given by

$$Z_{w,v} = \sqrt{\frac{\mu}{\kappa}} \cdot \frac{2a\lambda_g}{b\lambda} \dots \dots \dots (23)$$

Combining (19), (20), (22) and (23), and remembering that

$$(k^2 - h^2/k^2) = (\lambda/2b)^2,$$

we find

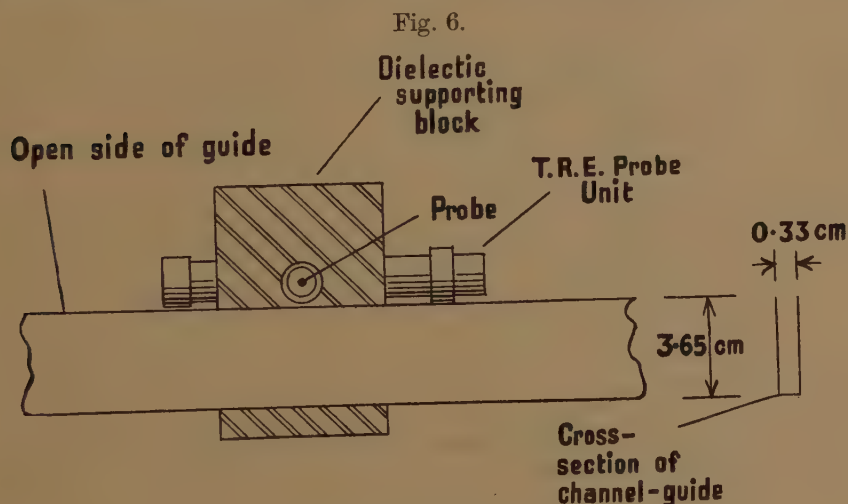
$$\alpha_r = \frac{\pi}{4} \cdot \frac{a\lambda_g}{b^3} \quad \dots \dots \dots (24)$$

The total attenuation coefficient is the sum of (21) and (24).

§ 5. EXPERIMENTAL RESULTS.

5.1. *Experimental Method.*

A schematic diagram of the apparatus is given in fig. 6. A block of dielectric material is arranged to form a sliding carriage on the waveguide.



It carries a T.R.E. probe unit and so measures the field about 5 mm. (about one-twentieth of the wavelength used) away from the open side of the channel. It was found experimentally that the reaction of the probe on the input impedance of the guide was negligible.

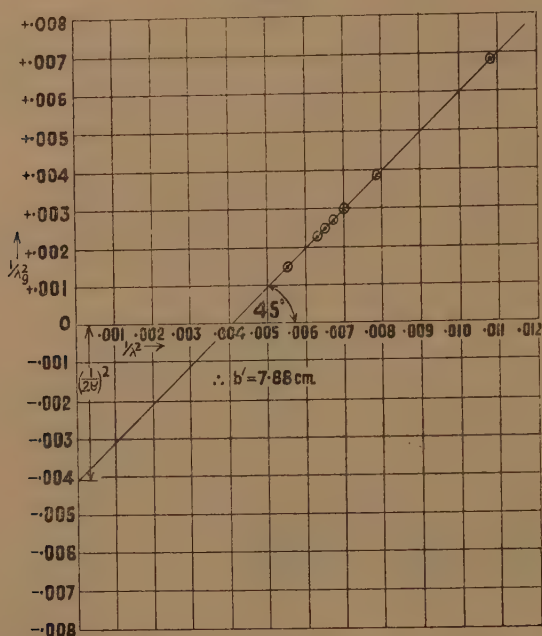
The reason for keeping the probe very close to the guide is to minimize end effect. The experimental procedure was to measure the rectified crystal current as a function of the position of the sliding carriage, at various wavelengths in the band 9–13.5 cm.

5.2. *Phase Constant.*

Measurements of the positions of minima were made, with a shorting plate at the end of the guide. Thus a direct measurement of guide wavelength could be obtained. In order to check the theory, $1/\lambda_g^2$ was plotted against $1/\lambda^2$, and this should give a straight line whose intercept on the $1/\lambda_g^2$ axis is $(1/2b')^2$. This graph is plotted in fig. 7. It will be

seen that the points lie very well on a straight line of unit slope. The intercept gives the effective width of the equivalent rectangular guide as 7.88 cm. Now $b=7.30$ cm. in the present case so that $\delta=0.58$ cm.

Fig. 7.



Experimental determination of b' , the effective breadth of the equivalent guide.

By substituting $b=7.30$ cm. and $a=0.33$ cm. in equation (8), we get $\delta=0.55$ cm., in reasonably good agreement with experiment.

5.3. Attenuation Constant.

As the crystal detector was operated on the square law part of its characteristic, the rectified current I was proportional to the power radiated in the immediate vicinity of the probe.

We can write

$$I = I_0 e^{2\alpha z},$$

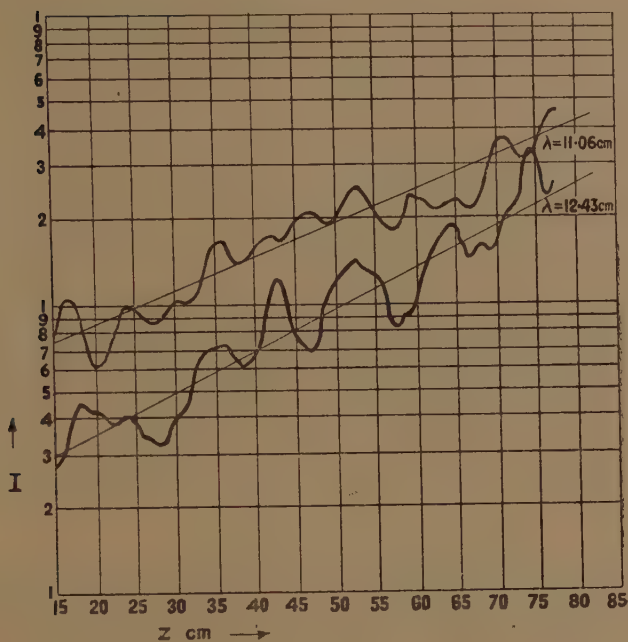
if z is measured from the termination end of the guide. Then

$$\log I = \log I_0 + 2\alpha z.$$

It is convenient to use log graph paper and to plot I on a logarithmic scale against z on a linear scale to get a straight line, from which α can be calculated.

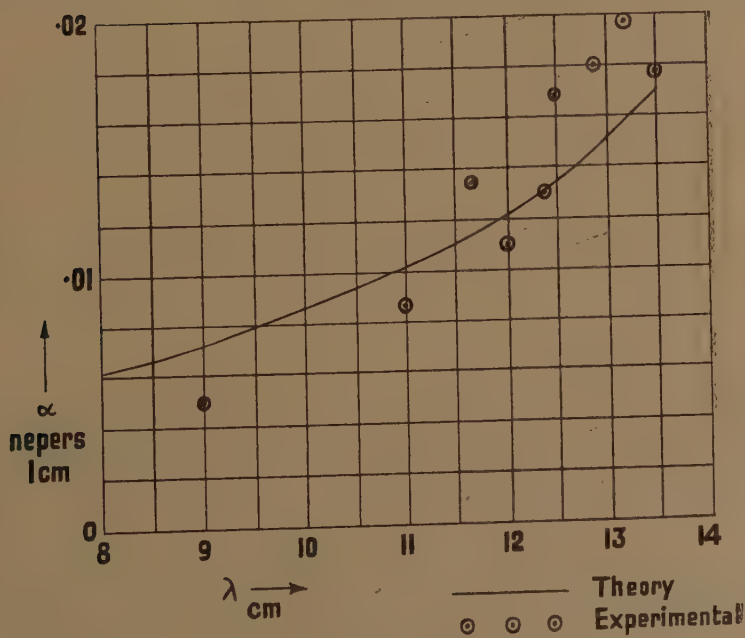
Typical experimental results are shown in fig. 8. The ripple is partly due to imperfect matching, and partly to mechanical imperfections. It is necessary to draw a mean straight line through each curve to find the appropriate value of α . Values of the attenuation constant α were then plotted as a function of wavelength, and the result is compared with the theoretical curve in fig. 9.

Fig. 8.



Distribution of radiated power along waveguide.

Fig. 9.



Attenuation constant.

The agreement of theory and experiment is thought to be satisfactory, for the theoretical curve is valid only if $\lambda \gg r \gg a$ and at $\lambda = 12$ cm., with $a = 0.33$ cm., the optimum choice of r gives $\lambda/r = r/a \sim 6$ which is certainly not $\gg 1$. A similar situation arises in the simple theory of a "thin" half-wave dipole.

§ 6. CONCLUSIONS.

The phase constant and attenuation constant of a channel section waveguide have been calculated, and simple approximate formulæ deduced. The validity of these formulæ has been verified by experiment, and satisfactory agreement has been found. The practical application of this device as a radiator is thought to be limited, for the attenuation rate is too great for really narrow beams to be obtained, and can only be reduced by reducing the open side dimension, which in turn would lead to increased wall losses and would reduce the power handling capacity. Moreover, the angle θ which the emergent beam makes with the normal to the array is rather large. For example, at $\lambda = 11$ cm. it is 45° , and at $\lambda = 13.5$ cm. it is 31° , which is still too large for most applications.

§ 7. ACKNOWLEDGMENTS.

Part of this paper has appeared as a Radio Department Report of R.A.E., Farnborough, and the author is indebted to the Chief Scientist, Ministry of Supply, for permission to publish it. Crown copyright is reserved. Reproduced with the permission of the Controller of H.M. Stationary Office. The author also wishes to acknowledge the encouragement he has received from Professor H. M. Barlow, of University College, London, who suggested the problem.

XXXVIII. *Regraduation in Spherically Symmetric Space-times of General Relativity.*

By T. J. WILLMORE, M.Sc., Ph.D.,
Durham Colleges*.

[Received March 30, 1948.]

§ 1.

THE possibility of extending to general relativity a theory of regraduation of clocks similar to that used in kinematic relativity, was examined by McVittie (1945) and by Walker (1946). These two papers differ fundamentally in their treatment of the problem, mainly in their different interpretation of what constitutes a regraduation. McVittie adopts a definition which implies that a regraduation is merely a coordinate transformation within space-time. Walker, on the other hand, treats a

* Communicated by Professor A. G. Walker, D.Sc.

regraduation as a local affair which changes the local coordinates but leaves unaltered the basic coordinates attached to an event. In this paper we adopt Walker's definition.

In this theory of regraduation, it appears that some features of a relativistic system are mathematical rather than physical, and these can be altered by a regraduation. We investigate whether spherical symmetry is such a feature, *i. e.* we consider whether an apparently non-symmetric space-time is transformable into a symmetric form by a regraduation. We show that such transformations cannot exist.

General conditions are then found for the existence of regraduations in models in which the pressure is isotropic and the metrics assume any spherically symmetric form. It is shown that if time-like regraduations are admitted, the models are necessarily of Lemaitre form. We then find the most general regraduation admitted by a Lemaitre model and by the Einstein static model, and conclude by showing that there is a class of spherically symmetric models which are merely equivalent to a Lemaitre model, their apparent non-homogeneity being mathematical rather than physical.

§ 2.

If X^a are local coordinates * and m represents mass at an event E, then a regraduation at E is a transformation of the form

$$\bar{X}^0 = uX^0, \quad \bar{X}^1 = vX^1, \quad \bar{m} = wm, \quad (1)$$

for some positive numbers u, v, w and all X^a, m . If values arising from this regraduation are denoted by a bar, then the energy and metrical tensors satisfy

$$\bar{T}^{ij} = wv^{-3}u^{-2}T^{ij}, \quad \bar{T}_{ij} = wv^{-3}u^2T_{ij}, \quad \bar{g}_{ij} = u^2g_{ij}. \quad . . (2)$$

If a regraduation takes place at every event E, then u, v, w exist for all x^i , and we have as equations relating the old and new tensor fields

$$\bar{T}_{ij} = e^{\psi}T_{ij}, \quad \bar{g}_{ij} = e^{2\sigma}g_{ij}, \quad (3)$$

where σ, ψ are functions of the x 's.

The original energy and metrical tensors are related by Einstein's field equations

$$-\kappa T_{ij} = G_{ij} - \frac{1}{2}Gg_{ij} + \Lambda g_{ij}, \quad (4)$$

where G_{ij} is the contracted curvature tensor, $G = g^{ij}G_{ij}$, and κ, Λ are constants ($\kappa \neq 0$).

If the regraduation is uniform, *i. e.* u, v, w are constants, it follows that if the original tensors satisfy (4), then the new tensors satisfy the same equation in barred quantities, the new constants $\bar{\kappa}, \bar{\Lambda}$ being given by the equations

$$\bar{\kappa} = \kappa e^{-\psi}, \quad \bar{\Lambda} = \Lambda e^{-2\sigma}. \quad (5)$$

* Roman suffixes will take values 0, 1, 2, 3; and Greek suffixes will take values 1, 2, 3.

Since equation (4) transforms into the field equations in the new system, it follows that every system admits all uniform regraduations. Walker has shown that the requirement that the path of a free particle be a geodesic restricts permissible regraduations to the uniform type. Following Walker, we shall relax this requirement.

A physical system together with a basic coordinate system, tensors g_{ij} , Γ_{ij} and constants κ, Λ satisfying (4) is called a model. Models M and \bar{M} are equivalent if their tensors satisfy equations of the form (3). It is assumed that the field equations are satisfied in each system, but it is not assumed that the constants $\bar{\kappa}, \bar{\Lambda}$ are related to κ, Λ in any definite way. If M, \bar{M} are equivalent, then each can be obtained from the other by a regraduation, and both may be regarded as different descriptions of the same physical system. Walker has shown that a given model can be regraduated if for some functions σ, ψ and some constants $\bar{\kappa}, \bar{\Lambda}$ the following conditions are satisfied :

$$(\kappa - \kappa e^\psi) \Gamma_{ij} = 2\sigma_{,ij} - 2\sigma_{,i} \sigma_{,j} - g_{ij} (2\Delta_2 \sigma + \Delta_1 \sigma + \Lambda - \bar{\Lambda} e^{2\sigma}), \quad . \quad . \quad (6)$$

where

$$\Delta_1 \sigma = g^{ij} \sigma_{,i} \sigma_{,j}, \quad \Delta_2 \sigma = g^{ij} \sigma_{,ij}, \quad . \quad . \quad . \quad (7)$$

and a comma denotes covariant differentiation.

§ 3.

In this section we show that equation (6) applied to spherically symmetric models restricts regraduations to the type $\sigma = \sigma(r, t)$, $\psi = \psi(r, t)$. From this it follows that spherical symmetry is preserved by regraduations.

The most general spherically symmetric form is

$$ds^2 = e^\nu dt^2 - e^\lambda dr^2 - e^\mu r^2 (d\theta^2 + \sin^2 \theta d\phi^2), \quad . \quad . \quad . \quad (8)$$

where $\lambda = \lambda(r, t)$, $\mu = \mu(r, t)$, $\nu = \nu(r, t)$. By a change of coordinate system it is possible to express (8) in either of the forms

$$ds^2 = e^\nu dt^2 - e^\lambda dr^2 - r^2 (d\theta^2 + \sin^2 \theta d\phi^2),$$

$$ds^2 = e^\nu dt^2 - e^\mu (dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2),$$

where the new spatial and temporal coordinates are functions of *both* the original spatial and temporal coordinates. In order to make our coordinates physically significant, we use the general form (8) and impose the condition that the gradient of t shall be a principal direction of the energy tensor, *i. e.* $T_{0\alpha} = 0$, $\alpha = 1, 2, 3$. This leads to the single relation between λ, μ, ν given by

$$\left(\frac{\partial \lambda}{\partial t} - \frac{\partial \mu}{\partial t} \right) \left(\frac{1}{r} + \frac{1}{2} \frac{\partial \mu}{\partial r} \right) + \frac{1}{2} \frac{\partial \mu}{\partial t} \frac{\partial \nu}{\partial r} - \frac{\partial^2 \mu}{\partial r \partial t} = 0. \quad . \quad . \quad (9)$$

Equation (6) with suffixes $i=0, j=1$ reduces to

$$\frac{\partial^2 \sigma}{\partial r \partial t} - \frac{\partial \sigma}{\partial r} \frac{\partial \sigma}{\partial t} - \frac{1}{2} \frac{\partial \lambda}{\partial t} \frac{\partial \sigma}{\partial r} - \frac{1}{2} \frac{\partial \nu}{\partial r} \frac{\partial \sigma}{\partial t} = 0. \quad . \quad . \quad . \quad (10)$$

The five other equations for which $i \neq j$ are found to have the general solution

$$e^{-\sigma} = r e^{\mu/2} \sin \theta \{F(\theta) + G(\phi)\} + H(r, t), \quad \dots \quad (11)$$

where F, G, H are arbitrary functions. If use is made of equation (9) it can be shown that (10) imposes no further restriction on F and G . However, it may be shown that the four remaining equations (6) for which $i=j$, are consistent with (11) only if $F(\theta) + G(\phi) = 0$. It follows that regraduations are restricted to the type $\sigma = \sigma(r, t)$, $\psi = \psi(r, t)$.

Equations (6) then reduce to (10) and the three equations

$$\begin{aligned} (\kappa - \bar{\kappa} e^{\psi}) T_0^0 = & -e^{-\nu} \left\{ 3 \left(\frac{\partial \sigma}{\partial t} \right)^2 + \frac{\partial \sigma}{\partial t} \left(\frac{\partial \lambda}{\partial t} + 2 \frac{\partial \mu}{\partial t} \right) \right\} \\ & + e^{-\lambda} \left\{ 2 \frac{\partial^2 \sigma}{\partial r^2} + \left(\frac{\partial \sigma}{\partial r} \right)^2 + \frac{\partial \sigma}{\partial r} \left(2 \frac{\partial \mu}{\partial r} - \frac{\partial \lambda}{\partial r} + \frac{4}{r} \right) \right\} - \Lambda + \bar{\Lambda} e^{2\sigma}, \end{aligned} \quad \dots \quad (12)$$

$$\begin{aligned} (\kappa - \bar{\kappa} e^{\psi}) T_1^1 = & -e^{-\nu} \left\{ 2 \frac{\partial^2 \sigma}{\partial t^2} + \left(\frac{\partial \sigma}{\partial t} \right)^2 + \frac{\partial \sigma}{\partial t} \left(2 \frac{\partial \mu}{\partial t} - \frac{\partial \nu}{\partial t} \right) \right\} \\ & + e^{-\lambda} \left\{ 3 \left(\frac{\partial \sigma}{\partial r} \right)^2 + \frac{\partial \sigma}{\partial r} \left(2 \frac{\partial \mu}{\partial r} + \frac{\partial \nu}{\partial r} + \frac{4}{r} \right) \right\} - \Lambda + \bar{\Lambda} e^{2\sigma}, \end{aligned} \quad \dots \quad (13)$$

$$\begin{aligned} (\kappa - \bar{\kappa} e^{\psi}) T_2^2 = & -e^{-\nu} \left\{ 2 \frac{\partial^2 \sigma}{\partial t^2} + \left(\frac{\partial \sigma}{\partial t} \right)^2 + \frac{\partial \sigma}{\partial t} \left(\frac{\partial \lambda}{\partial t} + \frac{\partial \mu}{\partial t} - \frac{\partial \nu}{\partial t} \right) \right\} \\ & + e^{-\lambda} \left\{ 2 \frac{\partial^2 \sigma}{\partial r^2} + \left(\frac{\partial \sigma}{\partial r} \right)^2 + \frac{\partial \sigma}{\partial r} \left(\frac{\partial \mu}{\partial r} + \frac{\partial \nu}{\partial r} - \frac{\partial \lambda}{\partial r} + \frac{2}{r} \right) \right\} - \Lambda + \bar{\Lambda} e^{2\sigma}. \end{aligned} \quad \dots \quad (14)$$

Thus we have shown that the model with general metric (8) can be regraduated if functions $\sigma(r, t)$, $\psi(r, t)$ are constants $\bar{\kappa}$, $\bar{\Lambda}$ can be found so that equations (9), (10), (12), (13), and (14) are satisfied.

§ 4.

We now restrict ourselves to models in which the pressure is isotropic, *i. e.* $T_1^1 = T_2^2 = T_3^3$. This imposes two additional relations, one obtained from the field equations (4), and the other from equations (13), (14). The first relation is

$$\begin{aligned} e^{-\lambda} \left\{ \frac{1}{2} \frac{\partial^2 \nu}{\partial r^2} - \frac{1}{2} \frac{\partial^2 \mu}{\partial r^2} - \frac{1}{2r} \frac{\partial \nu}{\partial r} - \frac{1}{2r} \frac{\partial \lambda}{\partial r} + \frac{1}{4} \left(\frac{\partial \nu}{\partial r} \right)^2 \right. \\ \left. - \frac{1}{r^2} - \frac{1}{4} \left(\frac{\partial \mu}{\partial r} \frac{\partial \nu}{\partial r} + \frac{\partial \nu}{\partial r} \frac{\partial \lambda}{\partial r} + \frac{\partial \lambda}{\partial r} \frac{\partial \mu}{\partial r} \right) \right\} + \frac{e^{-\mu}}{r^2} \\ + e^{-\nu} \left\{ \frac{1}{2} \frac{\partial^2 \mu}{\partial t^2} - \frac{1}{2} \frac{\partial^2 \lambda}{\partial t^2} + \frac{1}{2} \left(\frac{\partial \mu}{\partial t} \right)^2 - \frac{1}{4} \left(\frac{\partial \lambda}{\partial t} \right)^2 \right. \\ \left. - \frac{1}{4} \left(\frac{\partial \mu}{\partial t} \frac{\partial \nu}{\partial t} - \frac{\partial \nu}{\partial t} \frac{\partial \lambda}{\partial t} + \frac{\partial \lambda}{\partial t} \frac{\partial \mu}{\partial t} \right) \right\} = 0, \quad \dots \quad (15) \end{aligned}$$

and the second is

$$e^{-\lambda} \left\{ 2 \frac{\partial^2 \sigma}{\partial r^2} - 2 \left(\frac{\partial \sigma}{\partial r} \right)^2 - \frac{\partial \sigma}{\partial r} \left(\frac{\partial \lambda}{\partial r} + \frac{\partial \mu}{\partial r} + \frac{2}{r} \right) \right\} + e^{-\nu} \left\{ \left(\frac{\partial \mu}{\partial t} - \frac{\partial \lambda}{\partial t} \right) \frac{\partial \sigma}{\partial t} \right\} = 0. \quad (16)$$

The equations to be satisfied fall into two sets—equations (9), (15) which restrict the functions λ , μ , ν , and equations (10), (12), (13), (16) involving σ and ψ . The problem of finding the most general solution to these equations is not attempted in the present paper. We may simplify the problem by imposing restrictions on the function σ and find the most general metrics which admit these restricted regraduations: or we may further restrict the metric and find the most general regraduations admitted by these special metrics.

§ 5.

In this section we restrict σ to be a function of the temporal coordinate only. Using (9), (10) and (16), the metric assumes the form

$$ds^2 = e^\nu dt^2 - e^f \{ e^g dr^2 + e^h r^2 (d\theta^2 + \sin^2 \theta d\phi^2) \},$$

where ν, f are functions of t alone, and g, h are functions of r alone. There will be no loss of physical significance of the coordinates if we write

$$t_1 = \int e^{\nu/2} dt, \quad r_1 = r e^{h/2}.$$

It follows that a basic coordinate system can be found so that $\nu=0$, $e^f = R^2(t) \cdot m^2(r)$, $e^h = R^2(t)$. Equation (15) then restricts $m^2(r)$ to the form $(1 - kr^2)^{-1}$ where k is a constant which may be chosen to assume values 0, or ± 1 by absorbing the appropriate constant in $R(t)$. The metric thus reduces to the well-known Lemaitre form

$$ds^2 = dt^2 - R^2(t) \{ m^2(r) dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \}. \quad (17)$$

Thus we have shown that the only spherically symmetric systems in which the pressure is everywhere isotropic and which admit non-uniform regraduations of the restricted type $\sigma = \sigma(t)$, are those described by Lemaitre models.

Writing $T_0^0 = \rho$, $T_1^1 = -p$ in (12), (13) and eliminating ψ we get

$$2\rho\sigma'' + (\rho + 3p)\sigma'^2 + (4\rho + 6p)R'R^{-1}\sigma' + (\rho + p)(\bar{A} - \bar{A}e^{2\sigma}) = 0, \quad (18)$$

where dashes indicate differentiation with respect to t . This is the only restriction on σ . It follows that every Lemaitre model admits a non-uniform regraduation, and also that a regraduation can be found so that $\bar{\kappa}$ ($\neq 0$) and \bar{A} shall have prescribed values. These results are in agreement with Walker's paper (1946), except that his equation (28) corresponding to our (18) contains an erroneous factor 3 in the last member, evidently due to a misprint.

§ 6.

In this section we restrict the metrics to be of the Lemaitre type (17) and find the most general regradautions admitted by these models. Equations (9) and (15) are now satisfied identically. Equation (10) leads to

$$e^{-\sigma} = A(r) \cdot R(t) + B(t), \quad . \quad . \quad . \quad (19)$$

where A, B are arbitrary functions. Equation (16) gives

$$\frac{\partial^2 \sigma}{\partial r^2} - \left(\frac{1}{r} + \frac{1}{m} \frac{dm}{dr} \right) \frac{\partial \sigma}{\partial r} - \left(\frac{\partial \sigma}{\partial r} \right)^2 = 0. \quad . \quad . \quad . \quad (20)$$

Using (20), it can be shown that (19) becomes

$$\left. \begin{aligned} e^{-\sigma} &= \alpha(1 - kr^2)^{\frac{1}{2}} \cdot R(t) + x(t), & k &= \pm 1, \\ e^{-\sigma} &= \gamma r^2 R(t) + y(t), & k &= 0, \end{aligned} \right\}, \quad . \quad . \quad . \quad (21)$$

where α, a, γ, c are constants and $x(t), y(t)$ are so far undetermined.

Writing $T_0^0 = \rho, T_1^1 = -p$ in (12), (13) and eliminating ψ we get

$$\begin{aligned} 2\rho \frac{\partial^2 \sigma}{\partial t^2} + (\rho + 3p) \left(\frac{\partial \sigma}{\partial t} \right)^2 + (4\rho + 6p) R' R^{-1} \frac{\partial \sigma}{\partial t} \\ - m^{-2} R^{-2} \left\{ (4\rho + 6p) \frac{1}{r} \frac{\partial \sigma}{\partial r} + 3(\rho + p) \left(\frac{\partial \sigma}{\partial r} \right)^2 \right\} + (\rho + p) \{ \Lambda - \bar{\Lambda} e^{2\sigma} \} = 0. \end{aligned} \quad . \quad . \quad . \quad (22)$$

From (21), when $\alpha = 0$ or $\gamma = 0$ then $\sigma = \sigma(t)$ and the above equation reduces to (18); hence we shall assume $\alpha \neq 0, \gamma \neq 0$.

The field equations (4) reduce to

$$\left. \begin{aligned} \kappa \rho &= 3R^{-2}(R'^2 + k) - \Lambda, \\ \kappa p &= -R^{-2}(2RR'' + R'^2 + k) + \Lambda. \end{aligned} \right\}. \quad . \quad . \quad . \quad (23)$$

It is convenient to deal separately with the cases $k = \pm 1$ and $k = 0$.

Case 1. $k = \pm 1$.

Equations (21), (22), (23) give

$$(1 - kr^2)^{\frac{1}{2}} \cdot L(t) + M(t) = 0, \quad . \quad . \quad . \quad (24)$$

where

$$L(t) = 2\alpha\rho \{ (k + R'^2 - RR'')x - RR'x' + R^2x'' \}, \quad . \quad . \quad . \quad (25)$$

$$M(t) = (\rho + p)R \{ 3k\alpha^2 - 3x'^2 + \bar{\Lambda} - \Lambda x^2 \} + 2\rho Rxx'' + (4\rho + 6p)R'xx'. \quad (26)$$

Since (24) must be satisfied for all values of r , we must have $L(t) = 0, M(t) = 0$. These two conditions lead to the single equation

$$3k\alpha^2 - 3ku^2 - 3R^2u'^2 + \bar{\Lambda} = 0, \quad . \quad . \quad . \quad (27)$$

where we have written $x(t) = R(t) \cdot u(t)$.

We conclude that Lemaitre models for which $k = \pm 1$ admit regradautions of the type $\sigma = \sigma(t)$ if σ satisfies (18); and also that the most general regradaution involving *both* r and t is

$$e^{-\sigma} = R(t) \{ \alpha(1 - kr^2)^{\frac{1}{2}} + u(t) \}, \quad . \quad . \quad . \quad (28)$$

where
$$\int R^{-1}(t) dt = \int \left(k\alpha^2 + \frac{\bar{A}}{3} - ku^2 \right)^{-\frac{1}{2}} du. \quad (29)$$

Case 2. $k=0$.

By using an argument similar to that in Case 1, it may be shown that Lemaitre models for which $k=0$ admit regraduations of the type $\sigma=\sigma(t)$ if σ satisfies (18); and also that the most general regraduation involving both r and t is

$$e^{-\sigma} = R(t) \{ \gamma r^2 + v(t) \}, \quad (30)$$

where
$$\int R^{-1}(t) dt = -\frac{1}{2\gamma} \left(\frac{\bar{A}}{3} - 4\gamma v \right)^{\frac{1}{2}}. \quad (31)$$

§ 7.

In this section we apply the previous results to the Einstein static model, for which $k=1$, and $R(t)$ is a constant R_e . Equation (23) reduces to

$$\kappa p = 3R_e^{-2} - A, \quad \kappa p = -R_e^{-2} + A, \quad (32)$$

showing that the pressure and density are both constant.

Equations (28), (29) restrict regraduations involving the radial co-ordinate r to the form

$$e^{-\sigma} = R_e \left\{ \alpha(1-r^2)^{\frac{1}{2}} + \left(\alpha^2 + \frac{\bar{A}}{3} \right) \sin \left(\frac{t}{R_e} + \epsilon \right) \right\}, \quad (33)$$

where $\bar{A} + 3\alpha^2 \geq 0$, and ϵ is an arbitrary constant.

When $\bar{A} + 3\alpha^2 = 0$, σ involves the radial coordinate only; hence, at first it appears that regraduations of the type $\sigma=\sigma(r)$ will be admitted. This, however, proves to be false; for in this case equations (12), (13) lead to a zero value of $\bar{\kappa}$ which is inadmissible. We deduce that the Einstein static model will not admit regraduations of the type $\sigma=\sigma(r)$.

When $\alpha^2 + (\bar{A}/3) > 0$, the function ψ found from (12), (13) involves both r and t . It follows from equation (3) that the new density $\bar{\rho}$ and the pressure \bar{p} involve both r and t , so that the regraduated model is neither static nor homogeneous. A similar argument shows that if any general Lemaitre model be regraduated using a σ function given by (28) or (30), the new model will not be homogeneous. We are thus led to the interesting conclusion that any spherically symmetric model whose metric is included in the form

$$ds^2 = e^{2\sigma} \left\{ dt^2 - R^2(t) \left(\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \right) \right\}, \quad . . . (34)$$

where σ is given by (28) or (30), is merely equivalent to a Lemaitre model, the apparent non-homogeneity being mathematical rather than physical.

I wish to express my thanks to Professor Walker for many helpful suggestions and criticisms.

REFERENCES.

- McVITTIE, G. C., 1945, *Proc. Roy. Soc. Edin. A*, **62**, 147-155.
WALKER, A. G., 1946, *Proc. Roy. Soc. Edin. A*, **62**, 164-174.

XXXIX. *Theory of the High Pressure Mercury Vapour and Cadmium Vapour Discharges.*

By V. J. FRANCIS, B.Sc., M.I.E.E., A.R.C.S., F.Inst.P.

(Communication from the Staff of the Research Laboratories of The General Electric Company, Limited, Wembley, England) *.

[Received November 9, 1948.]

1. INTRODUCTION.

IN two recent papers (Francis 1946 a and b) hereafter called A and B respectively, an approximate solution of the energy balance equation for the high pressure mercury vapour discharge was found which made possible the calculation of many of the electrical characteristics. The families of curves representing the relations between current, voltage gradient, pressure and other parameters of the discharge were obtained in a form involving two constants which should be calculable from the atomic properties of mercury vapour (Francis 1946 b). In view of the uncertainty of the values of the atomic parameters, however, the constants were obtained by comparing the calculated curves with measured characteristics of the discharge. In this way it was shown that the solution of the energy balance equation gave results which agreed with experiment at least as well as could be expected in view of the uncertainties in the parameters of mercury vapour, and of the approximations made in the solution of the equation. It was decided therefore to apply the same method to the calculation of the characteristics of the high pressure cadmium discharge. These characteristics are of practical and commercial interest (Francis and Stevens 1947, Bourne and Beeson 1947), but are difficult to obtain experimentally even over that part of the range where the information would be of practical interest.

In the early two papers, there were some unsatisfactory features. It was recognized that the calculations were not carried out over the extended range of operating conditions, with the small intervals of argument and with the accuracy which the degree of accord with experiment would seem to justify. There were in addition some numerical errors (Burgess 1948). These did not modify the conclusions reached in the papers, but it seemed worth while to extend the calculations over a much wider range of the parameters and to improve their accuracy in so doing.

It was in fact necessary to do this in order to produce the results for cadmium vapour.

* Communicated by the Author.

Classical kinetic theory would give a temperature factor for thermal conductivity of the form $T^{\frac{1}{2}}$. Elenbaas in his early papers used the factor $T^{\frac{1}{2}}$ and more recently (Elenbaas 1947) has used $T^{\frac{1}{2}}$. There is little evidence to indicate the exact form of the relation (1.1) in the conditions of the high pressure arc.

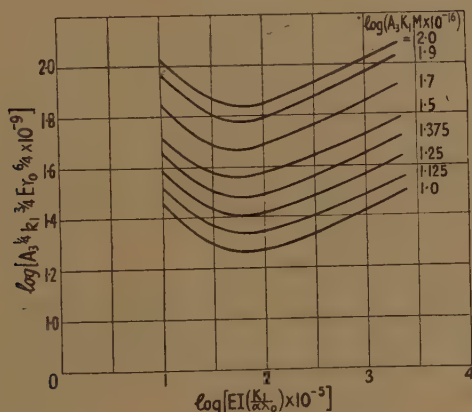
(c) It is assumed that radiation losses could be taken into account by the use of a single fictitious excitation level (taken as 8 volts for mercury) and that the excitation and ionization voltages are constant, although this is known to be unlikely at high pressures (Francis 1946 b, Rompe & Schultz 1938 and others).

These approximations were made for mercury in A and B and apply also in the present paper to both mercury and cadmium with, of course, appropriate changes in the values of the constants for the latter.

2. THE CALCULATIONS FOR MERCURY VAPOUR.

These were carried out in substantially the manner of paper B. A set of figures similar to that of fig. 1 of A and Table I. of B were, however,

Fig. 1.



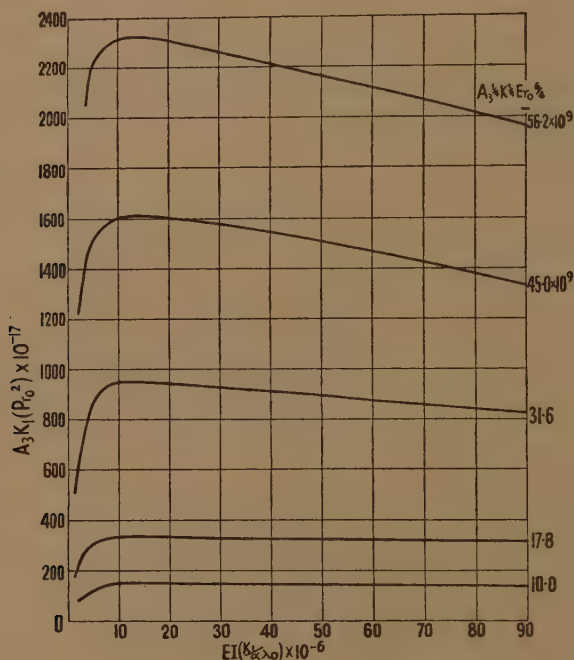
calculated for $A_3P^{\frac{1}{2}}/E^2 \times 10^{-8} = 12$ in addition to the values 2, 4, 6, 8, 10. The accuracy of the $\mu(T)$ values—which are most likely to introduce errors into the calculations—were improved by using five places of decimals, this accuracy being desirable because over an important part of the range, $\mu(T)$ is the difference of two nearly equal quantities. Smaller intervals in the argument $\beta/2T_w^2$ were also taken. The μ_0 values were calculated more carefully at the smaller values of $\beta/2T_w^2$ and EI , by extending the $\mu(T)$ curves back to values of T ranging from $T=4500^\circ$ for $A_3P^{\frac{1}{2}}/E^2 = 2 \times 10^8$ to $T=3000^\circ$ for $A_3P^{\frac{1}{2}}/E^2 = 12 \times 10^8$, and the errors (Burgess 1948) in the values of $S(x)$ were corrected. The use in A and B of a value of T_1 dependent only on $A_3P^{\frac{1}{2}}/E^2$ and not varying with $\beta/2T_w^2$ is unsatisfactory, since the ratio $\mu(T_1^2)/\mu(T_0^2)$ should be more or less constant for

all operating conditions, and have a magnitude of a few per cent if the approximation (a) in section 1 above is to be a good one. The values of T_0 calculated from equation (3.9) of A were taken as a first approximation to the axis temperature corresponding to the operating conditions specified by the values of $\beta/2T_0^2$ and $A_3P^{3/2}/E^2$. A second approximation to T_1 was then obtained by the relation

$$20\mu(T_1^2) = \mu(T_0^2), \quad (2.1)$$

except that when T_0 was on the right of the peak of the $\mu(T)$ curve, the peak value of μ was taken instead of the smaller value. The remainder of the calculations were as described in B.

Fig. 2.



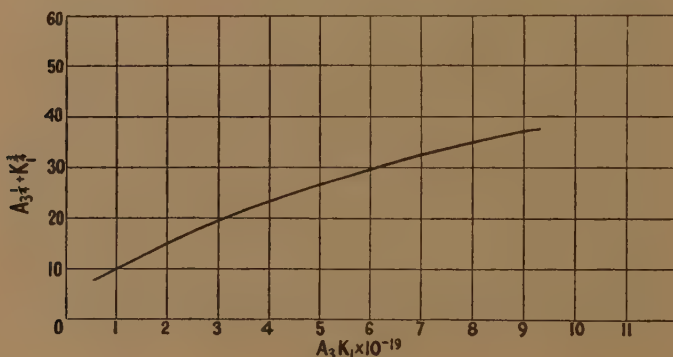
The relation between $Er_0^{3/2}$ and EI at constant M is shown over a wide range as a log-log plot in fig. 1. It will be seen that the absence of minima for these characteristics at small values of M indicated in fig. 3 of B, and thought there to result from the use of insufficiently small intervals in the arguments was, in fact, due to this cause. The minima occur at about the same value of EI for all values of M covered by the calculations, and the slope $d(\log Er_0^{3/2})/(d \log EI)$ of the "positive" part of the characteristic is substantially independent of M and EI . It follows that the proportional increase of voltage with increasing power input is independent of the arc voltage (or the mercury quantity) and of the loading.

This corresponds to the curves Y_1 and Y_2 in fig. 8 of B. Now if we attempt to obtain another relation between these constants from Elenbaas' data

$$\left. \begin{aligned} P &= 0.95 \\ EI &= 43.8 \times 10^7 \\ E &= 8.0 \times 10^8 \\ r_0 &= 1 \\ M &= 0.00306 \end{aligned} \right\}, \dots \dots \dots (2.3)$$

in equation (8.2) of B, and fig. 2 of the present paper, we obtain an almost identical curve, so that this gives no indication of the values required. A somewhat similar difficulty was found in B, but the increased accuracy of the present paper leads to the conclusion that the experimental data

Fig. 4.



so far used can be satisfied accurately by a wide range of the constants $A_3 K_1$ and $A_3^{1/2} K_1^{3/2}$. It is not immediately obvious why this should be so.

It is not possible to use the minima or rate of slope of the Er_0^3 vs. EI curves since, in contrast with what seemed likely from the calculations in B, there exist minima for all values of M , and moreover the proportional rate of increase of slope seems more or less dependent of M .

To resolve this difficulty, further experimental data are required. If we take the axial temperature of about 6000°K . estimated by Elenbaas for the conditions of equation (2.3) above, and use the curves similar to those of fig. 1 of B and the value of $K_1/\alpha\lambda_0$ in equation (2.2) we obtain at once

$$A_3^{1/2} K_1^{3/2} = 35. \dots \dots \dots (2.4)$$

It is difficult to assess how far the assumed value of $T_0 = 6000^\circ \text{K}$. is likely to be in error, but this is probably as good an estimate as can be made at present.

(N being Avogadro's Number and m_e the electron mass). So that

$$\frac{A'_3}{A_3} = 2 \frac{A'_1}{A_1} \dots \dots \dots (3.6)$$

For any particular line

$$A_1 \propto \nu^3 f, \dots \dots \dots (3.7)$$

where ν is the frequency of the line indicated and f is the oscillator strength. It is likely that the effective f -values for cadmium and mercury are very similar. Formally ν should correspond to the frequency of the single line which has been assumed to account for all the radiation. It is perhaps better, however, to take the resonance potentials as a relative measure of the energies involved in the transitions to which the radiation is actually due. We then obtain

$$\frac{\nu'^3}{\nu^3} = \left(\frac{3.8}{4.7} \right)^3 = 0.53, \dots \dots \dots (3.8)$$

which gives from (3.6) and (3.7)

$$A'_3 = 1.06 A_3 \dots \dots \dots (3.9)$$

and so

$$\left. \begin{aligned} A'_3 K'_1 &= 0.8 A_3 K_1 \\ A_3'^{\frac{1}{2}} K_1'^{\frac{1}{2}} &= 0.8 A_3^{\frac{1}{2}} K_1^{\frac{1}{2}} \end{aligned} \right\} \dots \dots \dots (3.10)$$

Finally, equations (2.2) and (2.9) give

$$\left. \begin{aligned} K'_1 / \alpha \lambda'_0 &= 1.5/43 = 0.035 \\ A'_3 K'_1 &= 5.12 \times 10^{19} \\ A_3'^{\frac{1}{2}} K_1'^{\frac{1}{2}} &= 24.8 \end{aligned} \right\} \dots \dots \dots (3.11)$$

4. INDEPENDENT CALCULATIONS FOR CADMIUM VAPOUR.

These followed closely the method used for mercury vapour. The law

$$\sigma(T)' = \sigma'_0 T \dots \dots \dots (4.1)$$

was used for the conductivity. The equations for the temperature

$$\frac{1}{r} \frac{d}{dr} \left\{ r \frac{dT^2}{dr} \right\} = -K'_1 E^2 P^{-\frac{1}{2}} \left[\lambda(T)' - \frac{A'_3 P^{\frac{1}{2}}}{E^2} \phi(T)' \right] \dots \dots (4.2)$$

holds where, however, K'_1 and A_3 are constants different from K_1 and A_3 for mercury vapour. The values

$$\left. \begin{aligned} V'_i &= 8.96 \\ V'_m &= 6.6 \end{aligned} \right\}, \dots \dots \dots (4.3)$$

were used for the ionization potential and the potential of the fictitious single excitation level. The functions

$$\left. \begin{aligned} \lambda(T)' &= T^{\frac{1}{2}} 10^{-\frac{2.26 \times 10^4}{T}} \\ \phi(T)' &= \frac{1}{T} 10^{-\frac{3.32 \times 10^4}{T}} \end{aligned} \right\}, \quad (4.4)$$

were calculated as for mercury to give the functions $\mu(T)'$ and μ'_0 , and equations similar to those of (3.9) and (3.10) of A.

The relation for P corresponding to equation (5.5) of A involves the atomic mass. It is preferable therefore to use the variable M/m' where m' is the mass of the cadmium atom whence

$$\left. \begin{aligned} \frac{M}{m'} &= \frac{2\pi \times 1.01 \times 10^6 P}{kT_w} S(x) \\ &= 4.63 \times 10^{22} \cdot \frac{P r_0^2}{T_w} S(x) \end{aligned} \right\}, \quad (4.5)$$

analogous to (5.6) of A.

An equation similar to (6.7) of A, but with different values of the constants, holds for I.

Sets of characteristics relating the quantities E, EI, P, M/m' , T_0 were thus calculated to be compared with results obtained in the next section.

5. DERIVATION OF CADMIUM VAPOUR CHARACTERISTICS FROM THOSE FOR MERCURY VAPOUR.

For this purpose, the fictitious level for cadmium was obtained by reducing that for mercury vapour in the same ratio as the ionization potentials. This gives the value

$$V'_m = 6.88. \quad (5.1)$$

We now wish to obtain the solution of the equation

$$\frac{1}{r} \frac{d}{dr} \left\{ r \frac{dT^2}{dr} \right\} = -K'_1 E^2 P^{-\frac{1}{2}} \left[T^{\frac{1}{2}} 10^{-\frac{\gamma e V_i}{4.6 k T}} - \frac{A'_3 P^{\frac{1}{2}}}{E^2} \cdot \frac{1}{T} 10^{-\frac{\gamma e V_m}{2.3 k T}} \right]. \quad (5.2)$$

from that for equation (1.2) of B where $\gamma = 0.86$. It is clear that

$$\frac{1}{r} \frac{d}{dr} \left\{ r \frac{dT^2}{dr} \right\} = -K'_1 E^2 P^{-\frac{1}{2}} \gamma^{\frac{1}{2}} \mu \left(\frac{T}{\gamma}, \frac{A'_3 P^{\frac{1}{2}}}{\gamma^{\frac{1}{2}} E^2} \right), \quad . . . (5.3)$$

where $\mu(x, y)$ is the same function as that used for the mercury vapour calculations. Equations (1.6), (1.8) and (1.9) of B still hold, where however

$$\beta' = \frac{\gamma^{\frac{1}{2}} K'_1 E^2 P^{-\frac{1}{2}} \mu'_0 r_1^2}{2} \quad (5.4)$$

and

$$\mu'_0 = \frac{2\gamma^2}{\beta'} \int_{T_1^2/\gamma^2}^{T_0^2/\gamma^2} \xi \left(T^2, \frac{A'_3 P^{\frac{1}{2}}}{\gamma^{\frac{1}{2}} E^2} \right) dT^2, \quad (5.5)$$

the $\xi(x, y)$ function being the same as that used for mercury vapour, due account, however, being taken of the change of the parameter $A_3 P^{\frac{2}{3}}/E^2$. The only other change is in the equation for EI which becomes

$$EI \left(\frac{K_1}{\alpha \lambda_0'} \right) = 0.137 \cdot \gamma^2 \cdot \frac{g(T_0^2/\gamma^2)}{\mu_0'}, \quad . \quad . \quad . \quad . \quad (5.6)$$

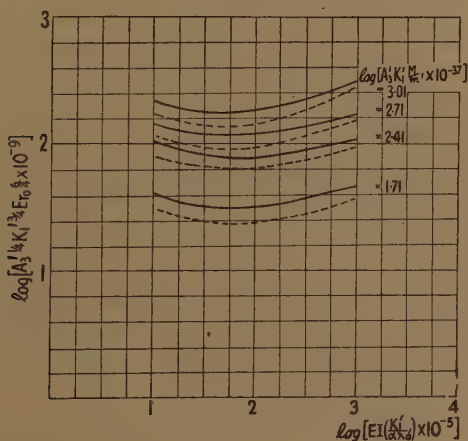
the g function also being the same as that used previously for mercury.

We may therefore use the tables for λ , ϕ and g , as calculated for mercury vapour, these now, however, referring to $A_3 P^{\frac{2}{3}}/\gamma^{\frac{2}{3}} E^2 = 2 \times 10^8$ etc. As an example of the method of tabulation adopted: we head a table with $A_3' P^{\frac{2}{3}}/\gamma^{\frac{2}{3}} E^2 = 4 \times 10^3$. This corresponds to Table I. of B with the modifications indicated in Section 2 of the present paper; we now list the same set of values of β' as was used for β in this Table; the values of T_1 and T_0 are the same as those obtained in Section 2 for mercury vapour, with $A_3 P^{\frac{2}{3}}/E^2 = 4 \times 10^8$; the value of μ_0' is given by (5.5), and in obtaining the (Pr_0^2) it is necessary to remember the γ factors both in (5.4) and in $A_3' P^{\frac{2}{3}}/\gamma^{\frac{2}{3}} E^2$. The relation (4.5) is used to obtain M/m' corresponding to column 9, Table I. of B and the other relations follow with the use of (5.6) instead of (1.12) of B. In this way without calculating the integral μ -curves for cadmium and using those already available for mercury, we obtain all the cadmium vapour characteristics.

6. COMPARISON OF THE TWO SETS OF CHARACTERISTICS FOR CADMIUM VAPOUR.

In Section 4 characteristics for cadmium vapour were obtained using the values of ionization and excitation potential of 8.96 and 6.6 volts

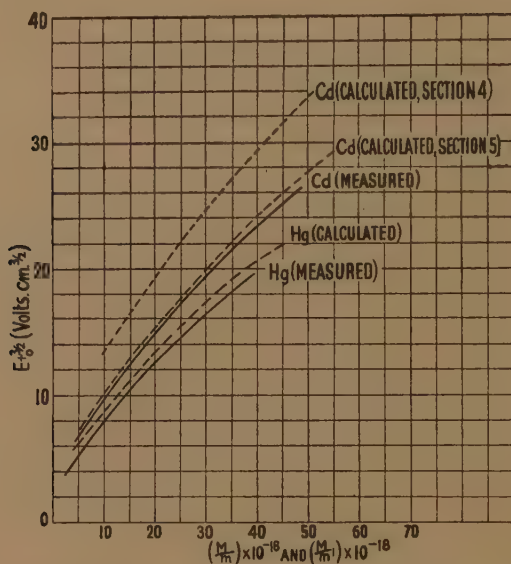
Fig. 5.



respectively. In Section 5 by a different method, characteristics were obtained for the values 8.96 and 6.88. The order of difference introduced by this change is shown in fig. 5, which gives the relation between E.

and EI at constant M/m' obtained by the first method (full lines) and the second method (dotted lines). It can be seen that the difference between 6.6 v. and 6.88 v. in the value for V'_m leads to a difference of the order of 15 per cent in the calculated characteristics. This difference is greater than errors due to other inaccuracies, which are probably of the order of 10 per cent, but is not so large as to suggest an unreasonable sensitivity of the results to small changes in V'_m . Fig. 6 shows the comparison between the results calculated in this paper and the few measurements available on cadmium vapour (Elenbaas 1937 b). The full lines give Elenbaas' measured values for mercury and for cadmium, the dotted

Fig. 6.



lines the values calculated here for mercury and for cadmium. In these calculations, the values of the constants for mercury obtained in Section 2 were used, and those for cadmium in Section 3.

The agreement between measured and calculated values is much better for the value of V'_m which has the same ratio to V'_i as that for the corresponding values for mercury vapour. For this reason the characteristics obtained using the value of $V'_m = 6.88$ are used in the following section.

7. COMPARISON OF ELECTRICAL CHARACTERISTICS FOR CADMIUM VAPOUR AND FOR MERCURY VAPOUR.

Using these calculations for cadmium, a comparison between the electrical characteristics for mercury and for cadmium may be made. Fig. 7, for example, shows the relation between E and M/m for mercury (dotted lines) and E and M/m' for cadmium (full lines) at constant EI . The

larger voltage gradient with cadmium compared with that with mercury is presumably due to the greater proportion of the input energy which is radiated from the cadmium discharge compared with that from the

Fig 7.

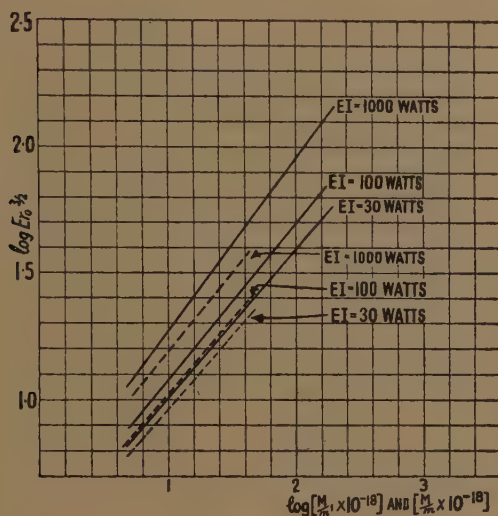
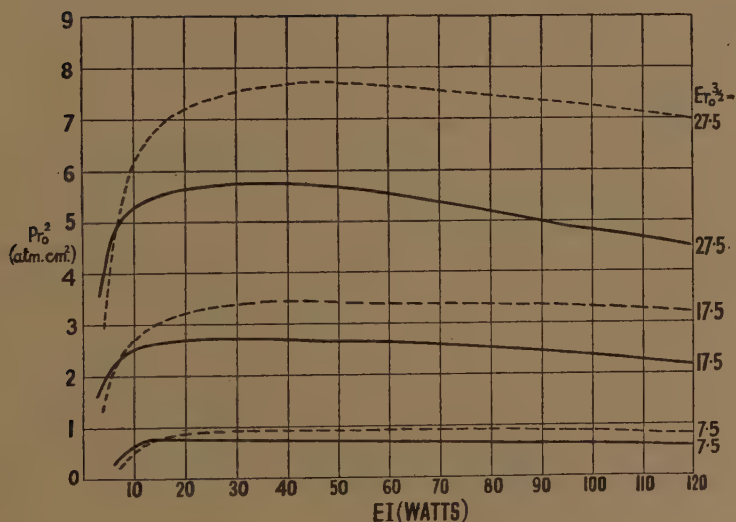


Fig. 8.

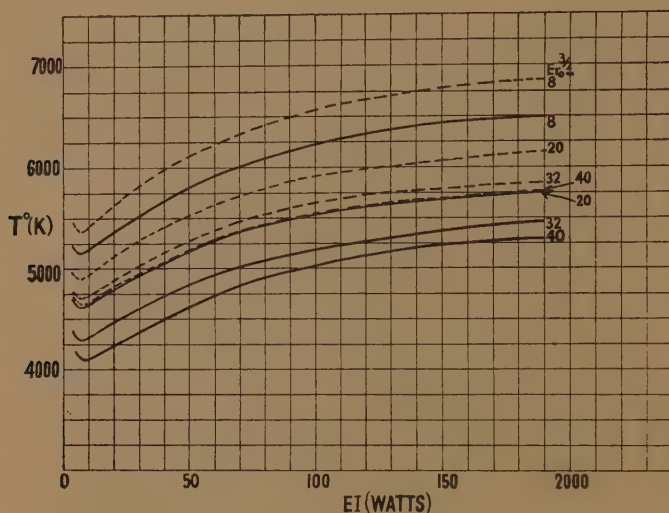


mercury vapour discharge resulting in a lower temperature and smaller degree of ionization.

Fig. 8 shows the pressure variation with varying loading at constant voltage gradient for both cadmium (full lines) and mercury (dotted lines),

and fig. 9 shows similarly the axial temperature as a function of voltage gradient and power input to the arc.

Fig. 9.



REFERENCES.

- BOURNE, H. K., and BEESON, E. J. G., 1947, *J. Brit. Kinema Soc.*, **11**, 107.
 BRAUNE, H., BASCH, R., and WENTZEL, 1928, *Z. Phys. Chem. (A)*, **137**, 447.
 BRODE, R. R., 1929, *Proc. Roy. Soc. A.*, **125**, 134; 1930, *Phys. Rev.*, **35**, 504.
 BURGESS, R. E., 1948, *Phil. Mag.* [7], **39**, 83.
 ELENBAAS, W., 1937 a, *Physica*, **4**, 279; 1937 b, *Physica*, **4**, 747; 1947, *Philips Tech. Rep.*, **2**, 20.
 FRANCIS, V. J., 1946a, *Phil. Mag.* [7], **37**, 433; 1946b, *Phil. Mag.* [7], **37**, 653.
 FRANCIS, V. J., and STEVENS, W. R., 1947, *J. Inst. El. Eng.*, **94**, 423.
 KILLIAN, T. J., 1930, *Phys. Rev.*, **35**, 1238.
 KLARFELD, B., 1937, *Tech. Phys. U.S.S.R.*, **4**, 44.
 MANNKOPF, R., 1936, *Z. f. Phys.*, **76**, 396.
 ROMPE, R., and SCHULTZ, P., 1938, *Z. f. Phys.*, **110**, 223.
 SCHULTZ, P., 1942, *Z. f. Phys.*, **119**, 167.

XL. *A Note on the Theory of Communication.*

By J. D. WESTON,
Sheffield University*.

[Received July 28th, 1948.]

§ 1.

THE purpose of this note is to suggest a basis for a general quantitative theory of communication. The fundamental idea is that a coded message can (like the state of a microphysical system in quantum theory) be represented as a vector, or point, in a space of an infinite number of dimensions, and that the process of transmitting the message over an "ideal" signalling system is (like a pure observation in quantum theory) equivalent to a projection of this vector on to a sub-space. Thus a message will be accurately transmitted if, and only if, it is so coded that its associated vector lies entirely in the sub-space characterizing the signalling system. A further idea is the provision, in this sub-space, of a set of rectangular axes with respect to which the components of a transmitted message are physical magnitudes associated with definite instants of time, the interval between consecutive instants being inversely proportional to the bandwidth of the signalling system. In this way the geometrical point of view is reconciled with that in which transmission is regarded as a process of interpolation.

§ 2.

It is necessary first to describe some results of which a fuller account appears elsewhere (Hardy 1941, Weston 1949). We denote by H_1 the Hilbert space determined by the functions of a real variable that are of summable square on the interval $(-\infty, \infty)$, and by H_3 the sub-space of H_1 consisting of the Paley-Weiner functions, that is, functions of the form †

$$\frac{1}{\sqrt{(2\pi)}} \int_{-\pi}^{\pi} F(u) e^{iut} du,$$

where $F(u)$ belongs to H_1 . H_3 is itself a Hilbert space, in which the functions $v_r(t)$ defined by

$$v_r(t) = \begin{cases} \frac{\sin \pi(t-r)}{\pi(t-r)} & \text{when } t \neq r, \\ 1 & \text{when } t = r \end{cases}$$

* Communicated by the Author.

† They may also be defined as continuous functions belonging to the class $L^2(-\infty, \infty)$ and having Fourier transforms that vanish almost everywhere outside the interval $(-\pi, \pi)$.

form a normal orthogonal basis for $r=0, \pm 1, \pm 2, \dots$. If $f(t)$ is a Paley-Wiener function, it can be expressed in the form

$$\sum_{r=-\infty}^{\infty} f(r)v_r(t)$$

(the convergence being uniform). This is known as a *cardinal series* and is familiar in the theory of interpolation (see, for example, Whittaker 1935). It is also a kind of Fourier series, and we have

$$f(r) = \int_{-\infty}^{\infty} f(t)v_r(t) dt,$$

and

$$\int_{-\infty}^{\infty} |f(t)|^2 dt = \sum_{r=-\infty}^{\infty} |f(r)|^2.$$

Since, with respect to the basis $\{v_r(t)\}$, the components of a Paley-Wiener function are its values at integral values of the variable, it is clear that if such a function vanishes at all the integers except n , it is a constant multiple of the function $v_n(t)$. The Fourier transform of this is the function $V_n(u)$ defined by

$$V_n(u) = \begin{cases} e^{-inu}/\sqrt{(2\pi)} & \text{when } |u| < \pi, \\ 0 & \text{when } |u| \geq \pi. \end{cases}$$

§ 3.

The communication of a message by a single-channel system consists essentially of three processes :—

- (i) some physical magnitude is caused to vary with time, the variations representing the message in accordance with some code, or language;
- (ii) these variations are propagated to a distant place (carrying energy with them);
- (iii) the variations are received and interpreted.

The “transmission system” is the physical apparatus in which the second of these processes takes place. As far as this system is concerned, a message is represented by a real function, $f(t)$, of a real variable (the time). We assume this function to be such that energy enters the system, at the sending end, at a rate proportional to $|f(t)|^2$ ($f(t)$ can always be chosen so that this is the case). The total energy imparted to the system by the “message” $f(t)$ is then proportional to $\int_{-\infty}^{\infty} |f(t)|^2 dt$. Since this must be finite for every message, we may regard a message as a point

in the space H_1 . We can then write

$$f(t) = \frac{1}{\sqrt{(2\pi)}} \int_{-\infty}^{\infty} F(u) e^{iut} du,$$

where $F(u)$, the Fourier transform of $f(t)$, belongs to H_1 , and the integral converges in mean square.

Now an actual system will not transmit $f(t)$ without distortion; transmission is accompanied by a transformation (bounded, because of the conservation of energy). An "ideal" transmission system may be defined as a system characterized by a transformation that is linear and that transforms e^{iut} into $e^{iu(t-\tau)}$ or zero according as $|u|$ is or is not in a certain interval (U_0, U) (τ is the delay in transmission, due to a finite velocity of propagation, and $(U - U_0)/2\pi$ is the bandwidth of the system). No generality is lost if we assume that $U_0 = 0$, since a fixed translation of frequency can be made at each end of the system with no resultant distortion (to make the appropriate change of frequency at the receiving end, it is not necessary to have advance knowledge of the message, but only of the method of transmission, and this may be regarded as part of the code, which must, of course, be known to the recipient). Moreover, we can choose the units in which time is measured in such a way that $U = \pi$. With this standardization of our ideal transmission system we have, as the associated transformation in H_1 , the projection, P , on the manifold H_3 .

It is evident that a necessary and sufficient condition for a message to be transmitted without distortion by the standard ideal system is that it lie entirely in H_3 . It is then the sum of a cardinal series. Moreover, since P is a singular transformation, any part of a message that is orthogonal to H_3 is irretrievably lost in transmission—nothing can be done at the receiving end to recover it.

§ 4.

The fact that if

$$f(t) = \frac{1}{\sqrt{(2\pi)}} \int_{-U}^U F(u) e^{iut} du$$

then, for any $k > 0$,

$$f\left(\frac{t}{k}\right) = \frac{k}{\sqrt{(2\pi)}} \int_{-U/k}^{U/k} F(ku) e^{iut} du,$$

may be expressed by stating that, for a given method of coding, the speed of signalling over an ideal transmission system is proportional to the bandwidth of the system. The bandwidth can therefore be regarded as a measure of the effectiveness of the system. It is usual, in fact, to take the bandwidth as a measure of the maximum possible speed of signalling over a system. However, to give an absolute meaning to the term "speed of signalling", it is necessary to define a unit signal. This has been done for special methods of coding, but not, apparently, in a completely general way.

In seeking to formulate a suitable definition, we may bear in mind the following requirements :—

- (i) a unit signal should be associated with a unique instant of time ;
- (ii) any message that can be transmitted without distortion over a given system should be a sum of unit signals, one for each of a succession of instants ;
- (iii) the unit signals composing a message should be independent of one another, in the sense that each can be perceived and interpreted without a knowledge of the others.

All these requirements are satisfied if, and only if, we define a unit signal to be a function that, when the time scale is suitably adjusted, is of the form $a_n v_n(t)$. The “unique instant” associated with this is the instant $t=n$, at which the function has the value a_n , and at which all its integral translations vanish (every function of the set $\{v_r(t)\}$ is an integral translation of the function $v_0(t)$, which vanishes at every integer except 0). The orthogonality of the functions $v_r(t)$ implies that the unit signals composing a message convey energy independently of one another.

With this definition, the maximum speed of signalling (number of unit signals per unit time) is unity for the standard ideal system. Of course, signalling could be carried on more slowly, but then the available bandwidth would not be used to full advantage.

It is clear from the form of the Fourier transform $V_n(u)$ that a signal approximating in form to $v_n(t)$ can be generated by applying a very short pulse to a good band-pass filter. Of course, the better the filter, the greater is the delay in transmission through it.

§ 5.

A unit signal, as just defined, has an amplitude which can be varied according to the meaning that is to be attached to the signal, and the reception of a transmitted message must now be regarded as, in effect, the observation of the amplitudes of the unit signals composing it. These are then interpreted in accordance with a known code. Ideally, any one of an indefinite number of meanings could be conveyed by a unit signal, each of a selected sequence of amplitudes representing a predetermined unit of language. In fact, however, owing to distortion and interference, each amplitude is to some extent indeterminate, and only a finite number of ranges of amplitude can be distinguished with reasonable certainty. This number is, evidently, proportional to the mean amplitude of the unit signals composing a typical message, that is, to the square root of the mean transmitted power. Since an observation necessarily involves the absorption of some energy from the system observed, one cannot measure an instantaneous amplitude, but only some kind of average over a short interval. This fact implies some degree of interference between different unit signals in the same message, and should be taken into account in estimating the effective signal-to-noise ratio of a communication channel.

The effects of distortion and interference can, of course, be analysed only by statistical means. It may be remarked, however, that the effect of a non-uniform frequency response in the transmission system is the production of "echoes"; that is to say, each unit signal is accompanied after transmission by a number of others, corresponding to different instants of time. The relative amplitudes of these are given by the coefficients in the complex Fourier series representing the frequency response of the system within its pass-band. Thus, in the case of the standard bandwidth, if the response of the system to a sine wave of pulsance u is given by the function

$$\sum_{r=-\infty}^{\infty} a_r e^{iur}$$

for $|u| \leq \pi$, then the unit signal $v_0(t)$ will become

$$\sum_{r=-\infty}^{\infty} a_r v_r(t)$$

when transmitted, as may be seen by considering Fourier transforms.

It may be thought that most methods of communication, telephony for example, differ very considerably from the "ideal" method that has been described here. However, it must be possible, in theory at any rate, to analyse every transmitted message as a cardinal series, since every transmission system has, for practical purposes, a finite bandwidth. The wastefulness of a given method of communication over an ideal system, free from interference, could be measured by the product of the time taken to transmit a given message and the bandwidth occupied by it, and direct telephony is a very wasteful method (this is shown by the economy in bandwidth that results from the use of the "vocoder"). Its principal advantage is that, when it is used for ordinary purposes, coding and interpretation are virtually instantaneous and do not require expensive apparatus.

I conclude by acknowledging my obligation to Mr. C. W. Earp, for many enlightening discussions on the theory and practice of communication.

REFERENCES.

- HARDY, G. H., 1941, *Proc. Camb. Phil. Soc.*, 37, 331-348.
 WESTON, J. D., 1949, *Proc. Camb. Phil. Soc.*, in the press.
 WHITTAKER, J. M., 1935, *Interpolatory Function Theory* (Cambridge: University Press).

XLI. *Stresses in Elastic Cylinders in Contact along a Generatrix
(including the effect of tangential friction).*

By EWEN M'EWEN,
University of Durham *.

[Received August 30th, 1948.]

§ 1. INTRODUCTION.

THE contact stresses in parallel elastic cylinders are usually treated as a special case of Hertz's classic solution for the contact stresses between any two bodies in contact whose dimensions and radii of curvature in the contact zone are both large compared with the contact area. Hertz's original papers (1881, 1882, 1896) give the general form of the solution, although he himself calculated only the normal stress component in the contact zone. Later papers (Baud 1931, Beeching and Nicholls 1948, Belajef 1917, 1924, Föppl 1936, Lundberg and Odqvist 1932, Prescott 1924, Thomas and Hoersch 1930) have completed the solution but have generally been similar in approach: Baud's approximate method (1931) is an exception but is numerically at variance with the others. None of these papers has included the effect of a tangential frictional component of load over the contact area, which is of importance in practical cases such as gears, cams, rollers, etc.

All these treatments assume that the contacting cylinders are relatively so large that the problem can be considered as that of two semi-infinite bodies in contact along a very narrow strip under conditions of plane strain. The same assumptions are made in the present paper, but we take into account tangential friction. Fig. 1 shows the external force on one of the contacting bodies: they are, of course, equal but reversed in sign on the other.

§ 2. NOTATION.

$2b$ = width of contact zone.

E = Young's modulus.

F = load per unit length of line of contact.

m, n = real functions (see equations (21) and (22)).

R_r = relative radius of curvature = $(R_1 R_2) / (R_1 + R_2)$.

R_1, R_2 = radii of contacting cylinders.

x, y, z = Coordinates. Ox is the line of contact, Oy is tangent to the contacting bodies and Oz the common normal to both bodies (fig. 1).

* Communicated by the Author.

\widehat{xx} , \widehat{yy} , \widehat{zz} =tensile stresses in directions Ox , Oy , Oz .

ξ , η , ζ =displacements in directions Ox , Oy , Oz .

γ =complex number ($=y+iz$).

Γ =complex function of γ ($=\phi+i\psi$).

$\Theta=(1-\sigma^2)/E$.

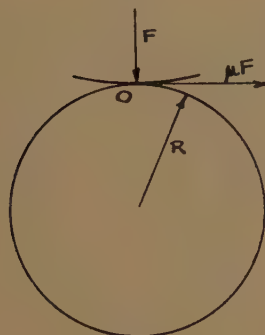
μ =coefficient of friction.

σ =Poisson's ratio.

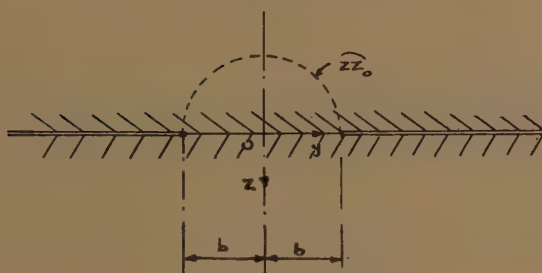
ϕ =real part of complex function Γ .

ψ =imaginary part of complex function Γ .

Fig. 1.



(a) External forces.



(b) The contact zone.

§ 3. ELASTIC EQUATIONS.

Since the problem is one of plane strain, the equations of equilibrium and compatibility are:—

$$\frac{\partial \widehat{yy}}{\partial y} + \frac{\partial \widehat{yz}}{\partial z} = \frac{\partial \widehat{yz}}{\partial y} + \frac{\partial \widehat{zz}}{\partial z} = 0, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\nabla^2(\widehat{yy} + \widehat{zz}) = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\xi = 0, \quad \widehat{xx} = -\sigma(\widehat{yy} + \widehat{zz}), \quad \widehat{xy} = \widehat{zx} = 0. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

§ 4. BOUNDARY CONDITIONS.

If the normal load F and tangential frictional load μF per unit length in the x direction are distributed over a strip extending from $y = -b$ to $y = +b$ in the boundary plane $z = 0$ of a semi-infinite elastic solid, we have,

in the contact zone :

$$z=0, \quad \int_{-b}^{+b} \widehat{zz} \, dy = -F, \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$\widehat{yz} = \mu \widehat{zz}. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Outside the contact zone :

$$\widehat{zz} = 0 \text{ when } z=0 \text{ and } |y| > b, \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$\left. \begin{array}{l} \widehat{yy} \rightarrow 0 \\ \widehat{zz} \rightarrow 0 \end{array} \right\} \text{ as } y \text{ or } z \text{ tend to infinity.} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

§ 5. GENERAL SOLUTION.

Let $\Gamma = \phi + i\psi$ be an analytic function of the complex variable $\gamma = y + iz$. Then the real and imaginary parts of Γ separately satisfy Laplace's equation in two dimensions and we have the Cauchy-Riemann relations

$$\frac{\partial \phi}{\partial y} = \frac{\partial \psi}{\partial z} \quad \text{and} \quad \frac{\partial \phi}{\partial z} = -\frac{\partial \psi}{\partial y}. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Write

$$\left. \begin{array}{l} \widehat{yy} = \phi + 2\mu\psi - z \frac{\partial \psi}{\partial y} + \mu z \frac{\partial \psi}{\partial z}, \\ \widehat{zz} = \phi + z \frac{\partial \psi}{\partial y} - \mu z \frac{\partial \psi}{\partial z}, \\ \widehat{yz} = \mu\phi - z \frac{\partial \psi}{\partial z} - \mu z \frac{\partial \psi}{\partial z}. \end{array} \right\} \quad . \quad . \quad . \quad . \quad . \quad (9)$$

We see immediately by differentiation that equations (9) satisfy the conditions of equilibrium and compatibility (1) and (2). Moreover, when $z=0$, $\widehat{yz} = \mu \widehat{zz}$ so satisfying the boundary condition (5).

§ 6. PARTICULAR SOLUTION FOR ELASTIC CYLINDERS IN CONTACT ALONG A GENERATRIX.

Write

$$\begin{aligned} \Gamma &= \frac{-2F}{\pi b^2} [\sqrt{b^2 - \gamma^2} + i\gamma], \\ &= \frac{-2F}{\pi b^2} [\sqrt{b^2 - y^2 + z^2 - i \cdot 2yz} + iy - z], \quad . \quad . \quad . \quad (10) \end{aligned}$$

then at $z=0$

$$\left. \begin{array}{l} \Gamma_{z=0} = \frac{-2F}{\pi b^2} [\sqrt{b^2 - y^2} + iy], \\ \phi_{z=0} = \frac{-2F}{\pi b^2} [\sqrt{b^2 - y^2}], \\ \psi_{z=0} = \frac{-2F}{\pi b^2} [y], \end{array} \right\} \quad . \quad . \quad . \quad . \quad (11)$$

which satisfy the boundary condition (6).

Further,

$$\widehat{z z}_{z=0} = \phi_{z=0} = \frac{-2F}{\pi b^2} [\sqrt{b^2 - y^2}]. \quad (12)$$

Hence

$$\int_{-b}^{+b} \widehat{z z}_{z=0} dy = \frac{-2F}{\pi b^2} \int_{-b}^{+b} \sqrt{b^2 - y^2} dy = -F, \quad (13)$$

which satisfies the total force boundary condition (4).

Finally

$$\Gamma = \frac{-2F}{\pi b^2} [\sqrt{b^2 - \gamma^2} + i\gamma] = \frac{-2F}{\pi b^2} i\gamma \left[1 - \left(1 - \frac{b^2}{2\gamma^2} + \dots \right) \right], \quad . . . (14)$$

which tends to zero as γ tends to infinity and so satisfies the boundary condition (7).

The chosen form for Γ therefore meets the boundary conditions. It remains to evaluate the constant b . Since the problem is one of plane strain

$$\begin{aligned} \frac{\partial \zeta}{\partial z} &= \frac{1}{E} \{ (1 - \sigma^2) \widehat{z z} - \sigma(1 + y) \widehat{y y} \}, \\ &= \frac{1 + \sigma}{E} \left\{ (1 - 2\sigma) \phi - 2\mu \sigma \psi - z \frac{\partial \phi}{\partial z} - \mu z \frac{\partial \psi}{\partial z} \right\}. \quad (15) \end{aligned}$$

Integrating with respect to z and differentiating twice with respect to y with the aid of the Cauchy-Riemann relations (8) we have

$$\frac{\partial^2 \zeta}{\partial y^2} = \frac{1 + \sigma}{E} \left\{ 2(1 - \sigma) \frac{\partial \psi}{\partial y} - \mu(1 - 2\sigma) \frac{\partial \phi}{\partial y} - z \frac{\partial^2 \phi}{\partial y^2} - \mu z \frac{\partial^2 \psi}{\partial y^2} \right\}. \quad (16)$$

In the contact zone $z=0$,

$$\frac{\partial \psi}{\partial y} = \frac{-2F}{\pi b^2} \quad \text{and} \quad \frac{\partial \phi}{\partial y} = \frac{+2F}{\pi b^2} \cdot \frac{y}{\sqrt{b^2 - y^2}},$$

$$\left. \frac{\partial^2 \zeta}{\partial y^2} \right|_{z=0} = \frac{-2F(1 + \sigma)}{\pi b^2 E} \left\{ 2(1 - \sigma) + \mu(1 - 2\sigma) \cdot \frac{y}{\sqrt{b^2 - y^2}} \right\}. \quad . . . (17)$$

Hence the changes in curvature of the two bodies are, recalling the skew-symmetry of the tangential friction

$$\left. \begin{aligned} - \frac{\partial^2 \zeta_1}{\partial y^2} \Big|_{z=0} &= \frac{2F(1 + \sigma_1)}{\pi b^2 E_1} \left\{ 2(1 - \sigma_1) + \mu(1 - 2\sigma_1) \cdot \frac{y}{\sqrt{b^2 - y^2}} \right\}, \\ - \frac{\partial^2 \zeta_2}{\partial y^2} \Big|_{z=0} &= \frac{2F(1 + \sigma_2)}{\pi b^2 E_2} \left\{ 2(1 - \sigma_2) - \mu(1 - 2\sigma_2) \cdot \frac{y}{\sqrt{b^2 - y^2}} \right\}. \end{aligned} \right\} \quad (18)$$

The relative radius of curvature before loading, R_r , is reduced by the load to zero. Hence the total change in curvature is equal to the original relative radius of curvature and we have

$$\frac{1}{R_r} = \frac{4F}{\pi b^2} \left\{ \Theta_1 + \Theta_2 + \frac{\mu}{2} \frac{y}{\sqrt{b^2 - y^2}} \left[\frac{(1 + \sigma_1)(1 - 2\sigma_1)}{E_1} - \frac{(1 + \sigma_2)(1 - 2\sigma_2)}{E_2} \right] \right\}, \quad \dots \dots (19)$$

where $\Theta = (1 - \sigma^2)/E$.

This expression is independent of y if either $\mu = 0$, that is, in the absence of a tangential frictional load, or if the bodies are of the same material. If the bodies are of different materials then in the presence of a tangential frictional load the stress distribution is asymmetrical about Oz and our assumed form of Γ no longer valid. The majority of practical cases, however, involve materials of the same elastic properties, generally hardened steel for both members.

Then either in the absence of tangential friction, or in its presence if materials of like elastic properties are involved,

$$b = \sqrt{\frac{4FR_r}{\pi} (\Theta_1 + \Theta_2)}, \quad \dots \dots (20)$$

which is Hertz's classic result.

§ 7. STRESS AT ANY POINT.

Let

$$\left. \begin{aligned} \phi &= \frac{-2F}{\pi b^2} [m - z], \\ \psi &= \frac{+2F}{\pi b^2} [n - y], \end{aligned} \right\} \dots \dots (21)$$

then, from equations (10) and (21)

$$\left. \begin{aligned} m &= \pm \sqrt{\frac{1}{2} \{ (b^2 - y^2 + z^2) + \sqrt{(b^2 - y^2 + z^2)^2 + 4y^2 z^2} \}}, \\ n &= \pm \sqrt{\frac{1}{2} \{ -(b^2 - y^2 + z^2) + \sqrt{(b^2 - y^2 + z^2)^2 + 4y^2 z^2} \}}, \end{aligned} \right\} \dots \dots (22)$$

where the sign of m is always the same as that of z and that of n is always the same as that of y .

Then

$$\left. \begin{aligned} \frac{\partial m}{\partial y} - \frac{\partial n}{\partial z} &= \frac{-y}{m} \cdot \frac{z^2 - m^2}{m^2 + n^2} = \frac{z}{n} \cdot \frac{y^2 - n^2}{m^2 + n^2}, \\ \frac{\partial m}{\partial z} = \frac{\partial n}{\partial y} &= \frac{z}{m} \cdot \frac{y^2 + m^2}{m^2 + n^2} = \frac{y}{n} \cdot \frac{z^2 + n^2}{m^2 + n^2}. \end{aligned} \right\} \dots \dots (23)$$

Substituting in equations (3) we obtain the following equations for the stress at any point:—

$$\left. \begin{aligned} \widehat{xx} &= \frac{+4F\sigma}{\pi b^2} [m - z + \mu(y - n)], \\ \widehat{yy} &= \frac{-2F}{\pi b^2} \left[m - 2z + 2\mu(y - n) + m \cdot \frac{z^2 + n^2}{m^2 + n^2} + \mu n \cdot \frac{z^2 - m^2}{m^2 + n^2} \right], \\ \widehat{zz} &= \frac{-2F}{\pi b^2} \left[m - m \cdot \frac{z^2 + n^2}{m^2 + n^2} - \mu n \cdot \frac{z^2 - m^2}{m^2 + n^2} \right], \\ \widehat{yz} &= \frac{-2F}{\pi b^2} \left[\mu(m - 2z) - n \cdot \frac{z^2 - m^2}{m^2 + n^2} + \mu m \cdot \frac{z^2 + n^2}{m^2 + n^2} \right]. \end{aligned} \right\} \dots \dots (24)$$

By writing $\mu=0$ in equations (24) we have the usual case without friction:—

$$\left. \begin{aligned} \widehat{xx}_{\mu=0} &= \frac{+4F\sigma}{\pi b^2} [m - z], \\ \widehat{yy}_{\mu=0} &= \frac{-2F}{\pi b^2} \left[m - 2z + m \cdot \frac{z^2 + n^2}{m^2 + n^2} \right], \\ \widehat{zz}_{\mu=0} &= \frac{-2F}{\pi b^2} \left[m - m \cdot \frac{z^2 + n^2}{m^2 + n^2} \right], \\ \widehat{yz}_{\mu=0} &= \frac{+2F}{\pi b^2} \left[n \cdot \frac{m^2 - z^2}{m^2 + n^2} \right]. \end{aligned} \right\} \dots \dots (25)$$

§ 8. CONCLUSION.

The results in the absence of friction agree with the classical solutions but have been here extended to the case of tangential friction for bodies of similar elastic properties only.

REFERENCES.

- BAUD, R. V., 1931, *Mechanical Engineering.*, **53**, 667.
 BEECHING, R., AND NICHOLLS, W., 1948, *Proc. Inst. Mech. E.*, **158**, 317.
 BELAJEF, N. M., 1917, *Bull. Inst. Engineers of Ways of Communication*, St. Petersburg. Quoted by TIMOSHENKO, S., 1934, *Theory of Elasticity* (New York: McGraw-Hill), p. 348; 1924, *Memoirs on the Theory of Structures*, (Leningrad: *Inst. Ways Comm.*). Quoted by TIMOSHENKO, S., *loc. cit.*
 FÖPPL, L., 1936, *Forsch. auf dem Gebiete des Ingenieurwesens*.
 HERTZ, H., 1881, *Jl. für die reine und angewandte Mathematik*, **92**, 156;
 1882, *Verhandlugen des Vereins zur Beförderung des Gewerbefleißes* (Nov.);
 1896, *Miscellaneous Papers* (London: Macmillan), p. 146 and 163.
 LUNDBERG, G., AND ODQVIST, F. K. G., 1932, *Proc. Ing. Vetenskaps Akad.*, No. 116.
 PRESCOTT, J., 1924, *Applied Elasticity* (London: Longmans), p. 638.
 THOMAS, H. R., AND HOERSCH, V. A., 1930, *Univ. Ill. Eng. Exp. Sta. Bul.*, No. 212.

XLII. *Determination of the Electron Energy Distribution in Gases from Townsend's Ionization Coefficient.*

By H. D. DEAS and K. G. EMELEUS,
Queen's University, Belfast*.

[Received November 12, 1948.]

1. INTRODUCTION.

WHEN a swarm of electrons of small charge density drifts through a gas at pressure p in a uniform field X , the number of new electrons produced by an electron in collisions per unit drift path, α , can be expressed (*e. g.* Townsend 1947) as

$$\alpha = pf(X/p). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

If the average drift velocity of the electrons at unit pressure in a field X is W , and the fraction of electrons per unit volume with energy of random motion between V and $V+dV$ is $F(V)dV$, then

$$\alpha/p = W^{-1} \int_I^\infty vF(V)P(V) dV, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where v is the speed of an electron with energy V , $P(V)$ the probability that an electron with energy V will ionize a molecule by collision in unit length of random path at unit pressure, and I is the ionization energy of the molecule (Emeleus, Lunt and Meek 1936). α/p , W , and the mean value of V ,

$$E = \int_0^\infty VF(V) dV, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

can be measured by experiments on electron swarms, and I and $P(V)$ can be measured by experiments on beams of electrons in gas at very low pressure. Comparison of the experimental values of α/p with those predicted from (2) for various assumed forms of $F(V)$ gives therefore a means of discriminating between the latter, although the use of experimental values for α/p , W , I , E and $P(V)$ in eqns. (2) and (3), and the normalizing equation

$$\int_0^\infty F(V) dV = 1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

cannot give a unique solution for $F(V)$. Proceeding in this way, Emeleus, Lunt and Meek (1936) have shown that for air, nitrogen, oxygen and hydrogen, $F(V)$ appears to have a form not far from the Maxwellian for values of X/p which vary from one gas to another, but which are in the range of a few tens to a few hundreds of volts/cm./ (mm. Hg).

* Communicated by Professor K. G. Emeleus.

In each gas, for lower values of X/p , there are considerable discrepancies between the experimental values of α/p and those calculated from (2) when $F(V)$ is assigned a Maxwellian form, possible reasons for which were discussed by Emeleus, Lunt and Meek. Subsequently, critiques of the experiments on swarms have been given by Loeb (1939), and by Druyvesteyn and Penning (1940), and some new experimental data have been obtained. Loeb has shown that use of the better experimental values of α/p now available for air and hydrogen necessitates no substantial change in the conclusions of Emeleus, Lunt and Meek. The question does, however, arise as to how sensitive their method is for discriminating between different forms of $F(V)$. In order to examine this point, calculations of α/p have been made for a form of $F(V)$ which results under certain conditions when the collisions between the electrons and molecules are elastic. In this case, the Maxwellian expression for $F(V)$,

$$cV^{\frac{1}{2}} \exp(-3V/2E) \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

is replaced by

$$CV^{\frac{1}{2}} \exp(-0.55V^2/E^2), \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

c and C being constants fixed by eqn. (4)*. The essential difference between (5) and (6) is that $F(V)$ decreases for high energy very much more rapidly for the latter than for the former. As a greater fraction of the electron collisions must be elastic when X/p is small than when it is large, it might be expected that use of (6) in (2) would then lead to better agreement with experiment than use of (5).

If energies are measured in electron-volts, W in cm./sec., X in volts/cm. and p in mm. Hg., and if

$$P(V) = b(V - I), \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where b is a constant, the expressions obtained for α/p from (2), by integrating and inserting numerical constants, are, if $F(V)$ is Maxwellian in form,

$$(\alpha/p)_M = 5.48 \cdot 10^7 b W^{-1} E^{\frac{1}{2}} (I + 4E/3) \exp(-3I/2E), \quad . \quad . \quad (8)$$

and if $F(V)$ has the form (6),

$$(\alpha/p)_{el} = 6.72 \cdot 10^7 b W^{-1} E^{\frac{1}{2}} \left(1 - 2\pi^{-\frac{1}{2}} \int_0^{y_i} \exp(-y^2) dy \right), \quad . \quad . \quad (9)$$

where $y_i = (I/E)0.55^{\frac{1}{2}}$.

The approximations made in using (7) and in neglecting $\frac{1}{2}mW^2$ relative to the mean energy of random electron motion, where m is the mass of an electron, neither of which is serious, have been used in the present work unless otherwise stated. In addition, the values used by Emeleus,

* Following usual practice, formula (6) will be referred to as the elastic distribution function. It can, however, also arise if the collisions between the electrons and molecules are inelastic, if the average fractional loss of energy of an electron in a collision is proportional to the square of the mean free path for that energy (Druyvesteyn and Penning 1940).

Lunt and Meek for E , I , W and b have again been used. No new values of E , I , and b are known, whilst in the one instance in which W has been redetermined, for pure nitrogen (Loeb 1939), no significant change results from use of the later values. No new calculations have been made for oxygen.

2. RESULTS OF CALCULATIONS : DIATOMIC GASES.

The results for air are shown in Table I., in which values of α/p calculated from (8) and (9) (columns 4 and 5 respectively), are compared with Sanders' experimental values $(\alpha/p)_{\text{exp}}$ (Loeb 1939). Values are also given in columns 4 and 5 for three values of E greater than the maximum in Sanders' experiments; the corresponding values of X/p and W have not been found by experiment, and the extrapolated values adopted by Emeleus, Lunt and Meek have again been employed.

TABLE I.
Values of α/p for air.

X/p	E	$(\alpha/p)_{\text{exp}}$		$(\alpha/p)_{\text{M}}$		$(\alpha/p)_{\text{el}}$	
20	2	3.4	10^{-5}	2.2	10^{-4}	6.7	10^{-17}
36	3	8.2	10^{-3}	1.1	10^{-2}	8.6	10^{-8}
55	4	9	10^{-2}	7.4	10^{-2}	1.5	10^{-4}
75	5	2.8	10^{-1}	2.4	10^{-1}	5.4	10^{-3}
100	6	6.4	10^{-1}	5.3	10^{-1}	4.1	10^{-2}
130	7	1.24		9.2	10^{-1}	1.5	10^{-1}
163	8	1.65		1.37		3.4	10^{-1}
239	10	—		2.45		1.04	
455	14	—		4.91		3.06	
595	16	—		6.14		4.19	

For values of X/p up to 163 the values of α/p calculated from the elastic distribution formula (6) diverge more from the experimental values than those calculated from the Maxwellian formula (5), $(\alpha/p)_{\text{el}}$ always having a smaller value than $(\alpha/p)_{\text{M}}$. For larger values of X/p and E , however, the ratio of the numbers in the last two columns becomes progressively less, although the relative magnitudes remain in the same order as for smaller values of X/p . This is a consequence of the fact that as E/I becomes large, the expressions (8) and (9) tend to asymptotic values of $7.3 \cdot 10^7 W^{-1/2} E^{3/2}$ and $6.7 \cdot 10^7 W^{-1/2} E^{3/2}$ respectively. The reason for the approach of the two values, as can be seen from (2), is that when E is of order of, or greater than I , a considerable fraction of the electrons are able to ionize at each collision, and since the spread of $F(V)$ round E is not greatly different in the two forms of $F(V)$, the resulting ionization is about the same. This will be true for any form of $F(V)$ which is not too sharply peaked. The consequences are twofold :—

(1) for such large values of X/p , more accurate experimental data are necessary to permit of discrimination between different assumed forms of $F(V)$ than are required for smaller values of X/p ;

(2) if it is desired to calculate α/p when this has not been found experimentally for large values of X/p , the choice of the form of $F(V)$ for use in (2) is less critical than for small values of X/p .

The results for hydrogen need not be described in the same detail, because the degree of agreement between values of α/p calculated from (2), using exact values of $P(V)$ and correcting for the energy of drift motion, and those obtained experimentally by Hale for mercury-free gas, has been discussed by Loeb (1939). The calculated values of α/p are a little greater than the experimental values for X/p greater than approximately 30. Table II. contains a selection of the values of $(\alpha/p)_M$ and $(\alpha/p)_{el}$ calculated from (8) and (9) for values of E which have been taken to be associated with the values of X/p and W used by Emeleus, Lunt and Meek.

TABLE II.
Values of α/p for hydrogen.

X/p	E	$(\alpha/p)_M$		$(\alpha/p)_{el}$	
13	2	4.6	10^{-4}	1.04	10^{-16}
31.5	4	1.1	10^{-1}	3.3	10^{-4}
57.5	6	4.9	10^{-1}	4.5	10^{-2}
88	8	9.6	10^{-1}	2.6	10^{-1}
123	10	1.42		6.3	10^{-1}
203	14	2.22		1.42	

TABLE III.
Values of α/p for nitrogen.

X/p	E	$(\alpha/p)_{exp}$ Posin	$(\alpha/p)_{exp}$ Bowls	$(\alpha/p)_M$	$(\alpha/p)_{el}$
20	2.20	8.7 10^{-5}	—	1.1 10^{-3}	3.9 10^{-13}
30	2.67	9.1 10^{-4}	—	6.2 10^{-3}	4.2 10^{-8}
50	4.0	3.3 10^{-2}	—	1.1 10^{-1}	3.3 10^{-1}
86	6.0	3.8 10^{-1}	3.3 10^{-1}	7.5 10^{-1}	6.9 10^{-2}
106	7.0	7.3 10^{-1}	5.1 10^{-1}	1.26	2.3 10^{-1}
160	9.0	1.79	1.20	2.56	9.3 10^{-1}

For hydrogen, the approach of the ratio of the two sets of calculated values of α/p to the asymptotic ratio becomes marked for smaller values of X/p than for air. This is because although the values of E are approximately the same for values of X/p up to about 10 in both gases, they increase more rapidly thereafter with increase of X/p for hydrogen than for air.

The results for nitrogen are shown in Table III. The two italicized values of E and the corresponding values of W , were obtained by extrapolation. Both Posin's values for nitrogen containing mercury vapour, and Bowls' values for mercury-free gas are given (Loeb 1939). Although

Posin's values do not relate to pure nitrogen, they are several orders of magnitude less than the experimental values available to Emeleus, Lunt and Meek, so that one of the discrepancies noted by these authors is removed. The differences between Posin's and Bowls' values are not serious for present purposes. The experimental values now lie between the values calculated from (8) and (9), as was the case for the upper range of X/p values in hydrogen.

The calculations for air, hydrogen and nitrogen make little alteration necessary to the conclusion of Emeleus, Lunt and Meek, that for the larger values of X/p , the still imperfectly determined experimental data for α/p , E and W are consistent with one another if $F(V)$ is Maxwellian. They also confirm that, so far as the experimental data are reliable, more serious discrepancies still exist for the lower values of X/p . The new point is, however, brought out that, particularly for the larger values of X/p , the Maxwellian distribution does not necessarily yield better results than the other distribution considered, and that with the present lack of accurate data for pure gases for all the quantities involved, comparatively large departures from a Maxwellian distribution could exist and pass undetected. This however does not detract from the value of calculations of the type considered, based on a Maxwellian distribution, for the larger values of X/p , in giving results likely to be accurate in order of magnitude, or in some instances better.

3. RESULTS OF CALCULATIONS: MONATOMIC GASES.

Emeleus, Lunt and Meek found that the values of α/p calculated for argon with the Maxwellian form of $F(V)$ were from 2 to 4 orders of magnitude greater than the experimental values then available, for X/p ranging from 2 to 15. This discrepancy is increased when later experimental values for pure argon are employed. Further, the disparity is not removed when the elastic distribution function (6) is used in place of the Maxwellian function (5), as this diminishes the calculated values of α/p by less than one order of magnitude for this range of values of X/p . New calculations have also shown that discrepancies of the same order occur with neon, for which reliable experimental values of α/p have been obtained (Druyvesteyn and Penning 1940). Evidently the rate of diminution of $F(V)$ with increase in V above about E is extremely rapid in these gases. In an attempt to obtain quantitative information on this point, a number of calculations have been made for neon, in which $F(V)$ has been assumed to have the form

$$A_n V^{\frac{1}{2}} \exp(-R_n V)^n, \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

with n ranging from 3 to 10. A_n and R_n can be found from eqns. (3) and (4). In terms of gamma functions,

$$R_n = \Gamma(5/2n) / E \Gamma(3/2n) \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and

$$A_n = n R_n^{\frac{3}{2}} / \Gamma(3/2n). \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

From (2) and (7), if e denotes the electronic charge, the expression for α/p is

$$\left(\frac{Ee}{150m}\right)^{\frac{1}{2}} \frac{Ib}{W} (\Gamma(5/2n)\Gamma(3/2n))^{-\frac{1}{2}} \{(R_n I)^{-1} \phi(3/n : x) - \phi(2/n : x)\}, \quad . \quad . \quad (13)$$

where $x = (R_n I)^n$ and

$$\phi(y : x) = \int_x^\infty t^{y-1} \exp(-t) dt. \quad . \quad . \quad . \quad (14)$$

The last expression is readily reduced to a form which can be evaluated from tables of the incomplete gamma functions. The expressions (8) and (9) are identical with (13) for the special cases $n=1$, and $n=2$.

Using Bailey's data (1924) for E and W as functions of X/p , and Kruithof and Penning's experimental values for α/p (Druyvesteyn and Penning 1940), it is found that agreement between theory and experiment cannot be obtained for a constant value of n . Thus for X/p equal to 1.7, the best agreement is obtained for a value of n between 9 and 10, and for X/p equal to 6.7, for n between 6 and 7. The position is unlikely to be altered radically by use of other values for E and W . The calculations do, however, emphasize the paucity of electrons with energies much in excess of the mean energy in these inert gases (Deas 1948).

SUMMARY.

A method for testing assumed forms of the energy distribution for electrons in gases by comparison of calculated and experimental value of Townsend's ionization coefficient α/p has been re-examined. It has been shown that no change is necessary in the previous conclusion that the experimental values of α/p are compatible, within the rather wide limits of experimental error, with the assumption that the energy distribution has a Maxwellian form for the larger values of X/p considered, in air, nitrogen and hydrogen, but that discrimination between different forms of the distribution function becomes increasingly less certain as X/p increases. New calculations for argon and neon again show that the distribution function falls off very rapidly with increase in energy above about the mean energy.

ACKNOWLEDGMENTS.

Use has been made in this work of calculations performed by the late Dr. T. McFadden (1942), whose results for the elastic distribution formula are in accord with ours. We are indebted to Dr. R. W. Lunt for many critical comments.

REFERENCES.

- BAILEY, 1924, *Phil. Mag.*, **47**, 379.
 DEAS, H. D., M.Sc., 1948, *Thesis*, Belfast.
 DRUYVESTEYN and PENNING, 1940, *Rev. Mod. Physics*, **12**, 87.
 EMELEUS, LUNT and MEEK, C. M., 1936, *Proc. Roy. Soc. A*, **156**, 394.
 LOEB, 1939, *Fundamental Processes of Electrical Discharge in Gases*, New York.
 MCFADDEN, T., 1942, *Ph.D. Thesis*, Belfast.
 TOWNSEND, J. S., 1947, *Electrons in Gases*, London.

XLIII. *On the Mechanical Forces in Dielectrics.*

By W. B. SMITH-WHITE,
University of Sydney *.

[Received June 1, 1948.]

SUMMARY.

Two distinct formulæ for the mechanical force acting in a dielectric are to be found in the literature of the subject. One formula due to Helmholtz is obtained by use of the principle of energy and the other is derived very simply from the Poisson-Kelvin theory of polarization. In this paper it is shown that the second formula is consistent with the principle of energy and that Helmholtz's theory is incompatible with the Poisson-Kelvin views on the cause of dielectric behaviour.

§ 1. INTRODUCTION.

THE mechanical force acting on the substance of a dielectric under electrical influence has been derived by Helmholtz (1881) by use of the principle of the conservation of energy. For a fluid of dielectric constant $K=1+4\pi k$ in an electric field \mathbf{E} he finds the force per unit volume

$$\mathbf{F}^{(h)} = -\frac{1}{2}\mathbf{E}^2\nabla k + \frac{1}{2}\nabla\left(\mathbf{E}^2\tau\frac{\partial k}{\partial\tau}\right) \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where τ is the density of the fluid and ∇ is the usual Hamiltonian symbol for gradient. More than fifty years ago Larmor (1897) criticized this energy method but later writers with the exception of Livens have considered that his criticisms "do not appear to be well founded" or else have overlooked them entirely. The Helmholtz theory is generally reproduced whenever the theory of the mechanical forces is discussed (Jeans 1927, Becker 1932). In this paper I support the view of Larmor that the Helmholtz energy method is fundamentally unsound.

The formula (1) is quite unlike any other fundamental formula in electrostatics and is restricted. The corresponding formulæ for a solid body (Pockels 1906) are very elaborate and unnatural. Further there are dielectric electrostatic systems, which we may call "hysteretic" systems, for which Helmholtz's theory can give no result at all. The method involves a process not used elsewhere in electrostatics, viz. the process of deforming the dielectric. Helmholtz equates the decrease in the electrostatic energy of the system to the work done by the mechanical forces in the deformation. However, in the development of the idea of an energy

* Communicated by Professor T. G. Room.

function in electrostatics there is nowhere any thought of deformable dielectrics. Thus the energy method involves the physical assumption that the electrostatic energy established on the basis of fixed dielectrics is still an energy function for the system when the dielectrics are regarded as deformable.

According to Pidduck (1925) (Livens 1926) "a satisfactory theory of dielectrics was first propounded by Kelvin on the lines of a corresponding theory of magnetization due to Poisson." This theory does not appear to be generally appreciated even by modern authors. It is shown that the behaviour of a dielectric under electrical influence can be explained by assuming that each volume element dv of dielectric acquires an electric moment $\mathbf{P}dv$. The vector \mathbf{P} is the polarization and the dielectric is said to have become polarized. Thus, for electrostatic purposes, we regard the dielectric as equivalent to a distribution of electric moment. Then the mechanical force on the volume element dv at a place where the field strength is \mathbf{E} is the force which would act on an electric dipole of strength $\mathbf{P}dv$ in the same circumstances. So we find the force per unit volume

$$\mathbf{F} = (\mathbf{P} \cdot \nabla) \mathbf{E}. \quad (2)$$

This formula is the only one consistent with Kelvin's theory of dielectrics and is general. For the liquid dielectric to which (1) applies $\mathbf{P} = k\mathbf{E}$ and (2) becomes

$$\mathbf{F} = \frac{1}{2} k \nabla E^2. \quad (3)$$

Then from (1) and (3)

$$\mathbf{F}^{(h)} - \mathbf{F} = \frac{1}{2} \nabla \left(E^2 \tau^2 \frac{\partial}{\partial \tau} \left(\frac{k}{\tau} \right) \right), \quad (4)$$

so that the two theories give different values for the mechanical force.

In adopting Kelvin's theory and with it the formula (2) two questions must be answered. Firstly we must show that (2) is in accord with the principle of energy and secondly we must show explicitly and in detail how it happens that Helmholtz's method leads to a wrong result. These questions are answered in this paper. Beginning with the formula (2) I calculate from it the work done by the mechanical forces in a deformation of the dielectric. It will then appear that in general a system consisting of a deformable dielectric body near an electric charge distribution is not mechanically conservative; there is no energy function in the sense assumed in Helmholtz's argument. Physical conservation of energy is achieved only by recognizing other energy changes which accompany a general variation of the system. The discussion depends primarily on a study of the deformation of a continuous medium which has associated with it a continuous density distribution. Though natural and simple these considerations do not appear to have been required for any other physical purpose and consequently have an element of novelty in them. Most of the formulæ in the next section are of course well known but have been expressed in notation which seems specially suited for our application.

§ 2. DEFORMATIONS OF CONTINUOUS DISTRIBUTIONS.

A deformation Δ of a continuous medium is determined when the displacement of every point or particle of the medium is specified. Taking a fixed rectangular system of axes, if y_1, y_2, y_3 be the coordinates of the particle which originally occupied the position x_1, x_2, x_3 then the deformation of a region is given by the three functions

$$y_i = y_i(x_1, x_2, x_3), \quad i = 1, 2, 3,$$

defined throughout this region. We suppose that the deformation can be achieved by a continuous movement from the initial to the final position; or, analytically, the deformation Δ is regarded as a member of a family of deformations $\Delta(t)$ which depend continuously on a parameter t :

$$y_i = y_i(x_1, x_2, x_3; t), \quad i = 1, 2, 3. \quad (5)$$

We take $t=0$ to correspond to zero deformation.

The Jacobian

$$v = v(t) = \frac{\partial(y_1, y_2, y_3)}{\partial(x_1, x_2, x_3)} \quad (6)$$

is positive.

An infinitesimal deformation is specified by the differentials

$$u_i = \frac{\partial y_i}{\partial t} \Delta t, \quad i = 1, 2, 3. \quad (7)$$

We shall be concerned especially with infinitesimal deformations from the initial undeformed state; the partial derivatives in (7) are then taken for the value $t=0$.

Consider a function of position in the medium; when the medium is deformed the value of the function associated with a particular point in space or with a particular particle of the medium will in general alter. In examining the variation of the function we may fix attention on a given point or on a given particle. For the deformation $\Delta(t)$ let

$$\phi^0 = \phi^0(x_1, x_2, x_3; t)$$

denote the value of the function at x_1, x_2, x_3 ; and let

$$\phi = \phi(x_1, x_2, x_3; t)$$

denote the value of the function at the position y_1, y_2, y_3 occupied by the particle which was originally at x_1, x_2, x_3 . Then, if y_1, y_2, y_3 are replaced by their values (5),

$$\phi(x_1, x_2, x_3; t) = \phi^0(y_1, y_2, y_3; t) \quad (8)$$

identically. In particular $\phi = \phi^0$ for $t=0$. Write

$$\delta\phi = \frac{\partial\phi^0}{\partial t} \Delta t, \quad (9)$$

$$\Delta\phi = \frac{\partial\phi}{\partial t} \Delta t. \quad (10)$$

Thus $\delta\phi$ is the differential variation of the function at a fixed point in space and $\Delta\phi$ is the variation associated with a particular particle. From (8), differentiating, and making use of the summation convention for a repeated suffix,

$$\frac{\partial\phi}{\partial t} = \frac{\partial\phi^0}{\partial y_\alpha} \cdot \frac{\partial y_\alpha}{\partial t} + \frac{\partial\phi^0}{\partial t}.$$

For $t=0$ this reduces to

$$\frac{\partial\phi}{\partial t} = \frac{\partial\phi}{\partial x_\alpha} \frac{\partial y_\alpha}{\partial t} + \frac{\partial\phi^0}{\partial t},$$

and so, multiplying by Δt , from (7), (9), (10)

$$\delta\phi = -\frac{\partial\phi}{\partial x_\alpha} u_\alpha + \Delta\phi. \quad (11)$$

For a vector function of position we have the corresponding formula

$$\delta\mathbf{E} = -\frac{\partial\mathbf{E}}{\partial x_\alpha} u_\alpha + \Delta\mathbf{E}, \quad (12)$$

obtained by applying (11) to each component of the vector separately.

Consider now a distribution of a physical quantity Q throughout the substance of a material medium. If the amount of Q belonging to a volume v of the medium is

$$\iiint_v P dx_1 dx_2 dx_3,$$

then P is the density of Q . A general variation of this system will involve a deformation of the medium and an alteration of the amount of Q belonging to the various parts of the medium. Fixing attention on a particular point in space we can distinguish three separate contributions to the total change in the density of Q at this place: the first due to a displacement of the medium, the second due to the dilatation of the medium and the third due to the proper change in the amount of Q belonging to the various parts of the medium.

Suppose the medium to have been subjected to the deformation $\Delta(t)$. Let P denote the density of Q and let p denote the corresponding "invariant" density, *i. e.* the amount of Q reckoned per unit of undeformed volume. Then

$$\iiint P dy_1 dy_2 dy_3 = \iiint p dx_1 dx_2 dx_3,$$

the integrals being taken over corresponding regions.

Hence

$$p = \nu P. \quad (13)$$

Differentiating,

$$\frac{\partial p}{\partial t} = \nu \frac{\partial P}{\partial t} + P \frac{\partial \nu}{\partial t};$$

then putting $t=0$, $\nu=1$ and multiplying by Δt

$$\Delta p = \Delta P + P \Delta \nu. \quad (14)$$

From (6)

$$\Delta v = \frac{\partial v}{\partial t} \Delta t = \Delta t \frac{\partial}{\partial t} \begin{vmatrix} \frac{\partial y_1}{\partial x_1}, & \frac{\partial y_1}{\partial x_2}, & \frac{\partial y_1}{\partial x_3} \\ \frac{\partial y_2}{\partial x_1}, & \frac{\partial y_2}{\partial x_2}, & \frac{\partial y_2}{\partial x_3} \\ \frac{\partial y_3}{\partial x_1}, & \frac{\partial y_3}{\partial x_2}, & \frac{\partial y_3}{\partial x_3} \end{vmatrix}$$

and, for $t=0$, $\partial y_i/\partial x_j=1$ or 0 as $i=j$ or $i\neq j$. Thus

$$\begin{aligned} \Delta v &= \Delta t \begin{vmatrix} \frac{\partial^2 y_1}{\partial x_1 \partial t}, & \frac{\partial^2 y_2}{\partial x_2 \partial t}, & \frac{\partial^2 y_1}{\partial x_3 \partial t} \\ 0, & 1, & 0 \\ 0, & 0, & 1 \end{vmatrix} + \dots + \dots \\ &= \frac{\partial}{\partial x_1} \left(\frac{\partial y_1}{\partial t} \Delta t \right) + \frac{\partial}{\partial x_2} \left(\frac{\partial y_2}{\partial t} \Delta t \right) + \frac{\partial}{\partial x_3} \left(\frac{\partial y_3}{\partial t} \Delta t \right) \\ &= \frac{\partial u_1}{\partial x_1} + \frac{\partial u_2}{\partial x_2} + \frac{\partial u_3}{\partial x_3} = \frac{\partial u_\alpha}{\partial x_\alpha} \dots \dots \dots (15) \end{aligned}$$

So, replacing ϕ by P in (11)

$$\begin{aligned} \delta P &= - \frac{\partial P}{\partial x_\alpha} u_\alpha + \Delta P \\ &= - \frac{\partial P}{\partial x_\alpha} u_\alpha - P \Delta v + \Delta p \\ &= - \frac{\partial P}{\partial x_\alpha} u_\alpha - P \frac{\partial u_\alpha}{\partial x_\alpha} + \Delta p. \end{aligned}$$

These three terms represent the contributions mentioned above. The result may be written

$$\delta P = - \frac{\partial}{\partial x_\alpha} (P u_\alpha) + \Delta p. \dots \dots \dots (16)$$

For a distribution of the vector quantity \mathbf{Q} with density \mathbf{P} the corresponding formula is

$$\delta \mathbf{P} = - \frac{\partial}{\partial x_\alpha} (\mathbf{P} u_\alpha) + \Delta \mathbf{p}. \dots \dots \dots (17)$$

We may call a physical quantity \mathbf{Q} “conservative” if the amount of \mathbf{Q} associated with any part of the medium is invariable. In the application to be made we shall use τ and ρ for the densities of such a conservative quantity. Then from (14) and (16)

$$\Delta \tau = - \tau \Delta v, \dots \dots \dots (18)$$

$$\delta \rho = - \frac{\partial}{\partial x_\alpha} (\rho u_\alpha). \dots \dots \dots (19)$$

§ 3. WORK DONE IN A DEFORMATION.

Consider a system consisting of a piece of dielectric substance under the influence of an electric charge distribution. When the charge is moved and the dielectric is deformed the mechanical forces acting on the parts of the system do work. We calculate this work for an infinitesimal variation of the system on the basis of formula (2).

Let us suppose that the electric charge consists of a continuous distribution with density ρ inside a volume v_1 bounded by a simple closed surface f_1 ; and that the dielectric occupies a volume v_2 bounded by a simple closed surface f_2 . The electrical state of the dielectric is specified by the polarization vector \mathbf{P} , which is the density of electric moment inside v_2 . The mechanical force acting on the charge in v_1 is $\rho\mathbf{E}$ per unit volume, and that acting on the dielectric in v_2 is $(\mathbf{P} \cdot \nabla)\mathbf{E}$ per unit volume. Thus in an infinitesimal deformation specified by the vector \mathbf{u} the work done by the mechanical forces is

$$\int_{v_1} \rho \mathbf{E} \cdot \mathbf{u} dv + \int_{v_2} ((\mathbf{P} \cdot \nabla)\mathbf{E}) \cdot \mathbf{u} dv.$$

Now

$$\rho \mathbf{E} \cdot \mathbf{u} = -\rho \frac{\partial \phi}{\partial x_\alpha} u_\alpha$$

where ϕ is the electric potential; and

$$((\mathbf{P} \cdot \nabla)\mathbf{E}) \cdot \mathbf{u} = \mathbf{P}_\beta \frac{\partial E_\alpha}{\partial x_\beta} u_\alpha = -\mathbf{P}_\beta \frac{\partial^2 \phi}{\partial x_\alpha \partial x_\beta} u_\alpha = +\mathbf{P}_\beta \frac{\partial E_\beta}{\partial x_\alpha} u_\alpha = \mathbf{P} \cdot \frac{\partial \mathbf{E}}{\partial x_\alpha} u_\alpha.$$

The above work is therefore

$$-\int_{v_1} \rho \frac{\partial \phi}{\partial x_\alpha} u_\alpha dv + \int_{v_2} \mathbf{P} \cdot \frac{\partial \mathbf{E}}{\partial x_\alpha} u_\alpha dv.$$

In addition to the volume forces just considered there is a traction on the surface f_2 bounding the dielectric. This traction is expressed by a suitable modification of the formula (2). Consider a dielectric medium of semi-infinite extent bounded by a plane surface; take the surface to be $x_3=0$ and suppose the dielectric to occupy the space $x_3 < 0$. The tangential components of electric force are continuous across the bounding surface but the normal component is discontinuous. We imagine the surface discontinuity in \mathbf{P} to be replaced by a thin layer in which \mathbf{P} changes rapidly but continuously, and regard the discontinuity as being generated when this layer becomes infinitely thin. As the limit is approached $\partial E_3 / \partial x_3$ becomes infinite but the other derivatives remain finite. The traction \mathbf{T} acting on the surface discontinuity may now be obtained by integrating the volume force throughout the layer covering unit area on the plane $x_3=0$. The components of force are

$$F_\alpha = \mathbf{P}_\beta \frac{\partial E_\alpha}{\partial x_\beta}$$

and the volume of integration tends to zero as we approach the limit so

that the only contribution to the traction is due to \mathbf{F}_3 ; thus $T_1=T_2=0$ and

$$T_3 = \int P_3 \frac{\partial E_3}{\partial x_3} dx_3.$$

The relation $\operatorname{div} \mathbf{E} = -4\pi \operatorname{div} \mathbf{P}$ reduces effectively to

$$\frac{\partial E_3}{\partial x_3} = -4\pi \frac{\partial P_3}{\partial x_3}.$$

Hence

$$T_3 = -4\pi \int P_3 \frac{\partial P_3}{\partial x_3} dx_3 = \left[-2\pi P_3^2 \right]_{P_3}^0 = 2\pi P_3^2.$$

As the normal component of the vector $\mathbf{D} = \mathbf{E} + 4\pi \mathbf{P}$ is continuous it follows that, if we distinguish the electric force intensities in $x_3 < 0$ and $x_3 > 0$ by the signs $-$ and $+$ respectively then $2\pi P_3 = \frac{1}{2}(E_{3+} - E_{3-})$ and so

$$T_3 = \frac{1}{2} P_3 (E_{3+} - E_{3-}).$$

Finally, since $\mathbf{E}_+ - \mathbf{E}_-$ is a vector perpendicular to the plane $x_3 = 0$, we may express this result in the form

$$\mathbf{T} = \frac{1}{2} \{ \mathbf{P} \cdot (\mathbf{E}_+ - \mathbf{E}_-) \} \mathbf{n}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

where \mathbf{n} is the unit normal from the $-$ to the $+$ side. This formula (20) applies generally to any free surface bounding a dielectric medium if the sign $-$ refers to inside the medium and the sign $+$ to outside the medium.

In the deformation considered above the work done by the traction on the surface f_2 is

$$\frac{1}{2} \int_{f_2} \{ \mathbf{P} \cdot (\mathbf{E}_+ - \mathbf{E}_-) \} \mathbf{n} \cdot \mathbf{u} df = \frac{1}{2} \int_{f_2} \{ \mathbf{P} \cdot (\mathbf{E}_+ - \mathbf{E}_-) \} n_\alpha u_\alpha df,$$

where \mathbf{n} is the unit normal from the $-$ to the $+$ side. Thus the total work done in the deformation by the mechanical forces acting in the system is

$$\Delta W = - \int_{v_1} \rho \frac{\partial \phi}{\partial x_\alpha} u_\alpha dv + \int_{v_2} \mathbf{P} \cdot \frac{\partial \mathbf{E}}{\partial x_\alpha} u_\alpha dv + \frac{1}{2} \int_{f_2} \mathbf{P} \cdot (\mathbf{E}_+ - \mathbf{E}_-) n_\alpha u_\alpha df.$$

We transform this result by Green's formula: let w_1, w_2, w_3 be three functions defined throughout a volume v bounded by a simple closed surface f , their partial derivatives of the first order being continuous inside and on to f . Then

$$\int_v \frac{\partial w_\alpha}{\partial x_\alpha} dv = \int_f n_\alpha w_\alpha df, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

where n_1, n_2, n_3 are the direction cosines of the outward drawn normal. The theorem is formally the same when v is the region lying outside one or more closed surfaces. Now

$$\begin{aligned} \int_{v_1} \rho \frac{\partial \phi}{\partial x_\alpha} u_\alpha dv &= \int_{v_1} \frac{\partial}{\partial x_\alpha} (\rho \phi u_\alpha) dv - \int_{v_1} \phi \frac{\partial}{\partial x_\alpha} (\rho u_\alpha) dv \\ &= \int_{f_1} \rho \phi n_\alpha u_\alpha df + \int_v \phi \delta \rho dv \end{aligned}$$

by (19). Also

$$\begin{aligned} \int_{v_2} \mathbf{P} \cdot \frac{\partial \mathbf{E}}{\partial x_\alpha} u_\alpha dv &= \int_{v_2} \frac{\partial}{\partial x_\alpha} (\mathbf{P} \cdot \mathbf{E} u_\alpha) dv - \int_{v_2} \mathbf{E} \cdot \frac{\partial}{\partial x_\alpha} (\mathbf{P} u_\alpha) dv \\ &= \int_{f_1} \mathbf{P} \cdot \mathbf{E}_- n_\alpha u_\alpha dv + \int_{v_2} \mathbf{E} \cdot \delta \mathbf{P} dv - \int_{v_2} \mathbf{E} \cdot \Delta \mathbf{p} dv \end{aligned}$$

by (17). Hence

$$\begin{aligned} \Delta W + \int_{v_2} \mathbf{E} \cdot \Delta \mathbf{p} dv &= - \int_{v_1} \phi \delta \rho dv - \int_{f_1} \rho \phi n_\alpha u_\alpha df + \int_{v_2} \mathbf{E} \cdot \delta \mathbf{P} dv \\ &\quad + \frac{1}{2} \int_{f_1} \mathbf{P} \cdot (\mathbf{E}_+ + \mathbf{E}_-) n_\alpha u_\alpha df. \quad \dots \dots \dots (22) \end{aligned}$$

It will appear in the next section that the right hand side of (22) is an exact differential. In fact if we write

$$V = \frac{1}{2} \int_{v_1} \rho \phi dv - \frac{1}{2} \int_{v_2} \mathbf{P} \cdot \mathbf{E} dv, \quad \dots \dots \dots (23)$$

the right hand side of (22) is $-\Delta V$. So we have, for the mechanical work done in an infinitesimal deformation

$$\Delta W = -\Delta V - \int_{v_2} \mathbf{E} \cdot \Delta \mathbf{p} dv, \quad \dots \dots \dots (24)$$

§ 4. GREEN'S RECIPROCAL THEOREM.

To prove the statement made at the end of the previous section we need the Green's reciprocal theorem. This theorem is a relation between two distributions of electric charge and electric moment. For our purpose consider a charge distribution with density ρ' in a volume v'_1 bounded by the surface f'_1 and suppose f'_1 surrounds f_1 ; also a moment distribution with density \mathbf{P}' in a volume v'_2 bounded by the surface f'_2 and suppose f'_2 surrounds f_2 . Let ϕ' , \mathbf{E}' and \mathbf{D}' refer to this second distribution. The vector \mathbf{D} is continuous everywhere except on f_2 and across this surface the normal component of \mathbf{D} is continuous. Applying Green's transformation (21) to each of the five compartments of space separated by f_1, f'_1, f_2, f'_2 with $w_\alpha = \phi' D_\alpha$ and adding the results we have

$$\int_S \frac{\partial}{\partial x_\alpha} (\phi' D_\alpha) dv = 0,$$

where \int_S signifies an integration throughout the whole of space. As $\text{div } \mathbf{D} = 4\pi\rho$ inside v_1 and $\text{div } \mathbf{D} = 0$ outside v_1 we have

$$\begin{aligned} \int_{v_1} \rho \phi' dv &= \frac{1}{4\pi} \int_{v_1} \phi' \text{div } \mathbf{D} dv = \frac{1}{4\pi} \int_S \phi' \text{div } \mathbf{D} dv \\ &= \frac{1}{4\pi} \int_S \phi' \frac{\partial D_\alpha}{\partial x_\alpha} dv = \frac{1}{4\pi} \int_S \frac{\partial}{\partial x_\alpha} (\phi' D_\alpha) dv - \frac{1}{4\pi} \int_S D_\alpha \frac{\partial \phi'}{\partial x_\alpha} dv \\ &= \frac{1}{4\pi} \int_S D_\alpha E'_\alpha dv = \frac{1}{4\pi} \int_S \mathbf{D} \cdot \mathbf{E}' dv. \quad \dots \dots \dots (25) \end{aligned}$$

Similarly

$$\int_{v'_1} \rho' \phi dv = \frac{1}{4\pi} \int_S \mathbf{D}' \cdot \mathbf{E} dv.$$

Hence

$$\begin{aligned} \int_{v_1} \rho \phi' dv - \int_{v'_1} \rho' \phi dv &= \frac{1}{4\pi} \int_S (\mathbf{D} \cdot \mathbf{E}' - \mathbf{D}' \cdot \mathbf{E}) dv \\ &= \int_S (\mathbf{P} \cdot \mathbf{E}' - \mathbf{P}' \cdot \mathbf{E}) dv \\ &= \int_{v_2} \mathbf{P} \cdot \mathbf{E}' dv - \int_{v'_2} \mathbf{P}' \cdot \mathbf{E} dv. \end{aligned}$$

This is the relation required.

If v'_1 and v'_2 be the parts of v_1 and v_2 lying outside v_1 and v_2 , this result may be written

$$\int_{v_1} (\rho \phi' - \rho' \phi) dv - \int_{v'_1} \rho' \phi dv = \int_{v_2} (\mathbf{P} \cdot \mathbf{E}' - \mathbf{P}' \cdot \mathbf{E}) dv - \int_{v'_2} \mathbf{P}' \cdot \mathbf{E} dv.$$

We specialize this formula to the case in which the second distribution is obtained by an infinitesimal variation of the first. Then

$$\rho \phi' - \rho' \phi = \rho(\phi' - \phi) - \phi(\rho' - \rho) \rightarrow \rho \delta \phi - \phi \delta \rho,$$

and

$$\mathbf{P} \cdot \mathbf{E}' - \mathbf{P}' \cdot \mathbf{E} = \mathbf{P} \cdot (\mathbf{E}' - \mathbf{E}) - \mathbf{E} \cdot (\mathbf{P}' - \mathbf{P}) \rightarrow \mathbf{P} \cdot \delta \mathbf{E} - \mathbf{E} \cdot \delta \mathbf{P}.$$

Also, if the infinitesimal translation \mathbf{u} moves f_1 to f'_1 and f_2 to f'_2 ,

$$\begin{aligned} \int_{v'_1} \rho' \phi dv &\rightarrow \int_{f_1} \rho \phi n_\alpha u_\alpha df, \\ \int_{v'_2} \mathbf{P}' \cdot \mathbf{E} dv &\rightarrow \int_{f_2} \mathbf{P} \cdot \mathbf{E}_+ n_\alpha u_\alpha df. \end{aligned}$$

We now have the formula

$$\int_{v_1} (\rho \delta \phi - \phi \delta \rho) dv - \int_{f_1} \rho \phi n_\alpha u_\alpha df = \int_{v_2} (\mathbf{P} \cdot \delta \mathbf{E} - \mathbf{E} \cdot \delta \mathbf{P}) dv - \int_{f_2} \mathbf{P} \cdot \mathbf{E}_+ n_\alpha u_\alpha df. \quad (26)$$

Though our derivation of (26) has assumed that the surfaces f_1 and f_2 have moved uniformly outward it will be clear that this formula is valid without any such restriction.

Now differentiating (23) we have

$$\begin{aligned} \Delta V &= \frac{1}{2} \int_{v_1} (\rho \delta \phi + \phi \delta \rho) dv + \frac{1}{2} \int_{f_1} \rho \phi n_\alpha u_\alpha df \\ &\quad - \frac{1}{2} \int_{v_2} (\mathbf{P} \cdot \delta \mathbf{E} + \mathbf{E} \cdot \delta \mathbf{P}) dv - \frac{1}{2} \int_{f_2} \mathbf{P} \cdot \mathbf{E}_- n_\alpha u_\alpha df \\ &= \int_{v_1} \phi \delta \rho dv + \int_{f_1} \rho \phi n_\alpha u_\alpha df - \int_{v_2} \mathbf{E} \cdot \delta \mathbf{P} dv - \frac{1}{2} \int_{f_2} \mathbf{P} \cdot (\mathbf{E}_+ + \mathbf{E}_-) n_\alpha u_\alpha df \end{aligned}$$

by (26). This completes the proof of (24).

§ 5. COMPARISON WITH HELMHOLTZ'S THEORY.

With Larmor we shall regard the function V as the purely electrical energy of the system. Specializing (25) for the case when the two distributions are identical we have

$$\frac{1}{2} \int_{v_1} \rho \phi dv = \frac{1}{8\pi} \int_S \mathbf{D} \cdot \mathbf{E} dv$$

and, since $\mathbf{P}=0$ outside v_2 ,

$$\frac{1}{2} \int_{v_1} \mathbf{P} \cdot \mathbf{E} dv = \frac{1}{2} \int_S \mathbf{P} \cdot \mathbf{E} dv.$$

Hence, by subtraction

$$V = \frac{1}{8\pi} \int \mathbf{E}^2 dv. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

We may then suppose this electrical energy distributed throughout space with a density $\mathbf{E}^2/8\pi$.

The function V is not in general a mechanical potential function for the system. In the special case, however, of a hypothetical permanently polarized dielectric in which electric moment is a conservative quantity in the sense of § 2, then $\Delta \mathbf{p}=0$ and (24) reduces to $\Delta W = -\Delta V$, so that V is a mechanical potential function for this system.

The general formula (24) is new and its formal derivation from (2) is the significant development of this paper. In special cases it is equivalent to known formulæ. Thus if we regard the dielectric as non-deformable $\Delta \mathbf{p}=\delta \mathbf{P}$ and then from (24) and (27)

$$\begin{aligned} \Delta W &= -\Delta V - \int_{v_1} \mathbf{E} \cdot \delta \mathbf{P} dv \\ &= -\frac{1}{4\pi} \int_S \mathbf{E} \cdot \delta \mathbf{E} dv - \int_S \mathbf{E} \cdot \delta \mathbf{P} dv \\ &= -\frac{1}{4\pi} \int_S \mathbf{E} \cdot \delta \mathbf{D} dv, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28) \end{aligned}$$

which is a known result. If further the dielectrics in the system are "non-hysteretic" the expression on the right is a perfect differential and the system is mechanically conservative. The corresponding integral is the electrostatic energy of the system.

The Helmholtz theory of the mechanical forces in dielectrics sets out from the formula (28) derived independently of the theory of polarization of dielectrics and assumes the formula is valid when the dielectrics in the system are deformed. The theory itself is self-consistent; it is incompatible with Kelvin's view on the cause of dielectric behaviour, which involves the formula (2). The difference between the two theories appears in the significance attached to (28). Helmholtz regards this formula as true generally whereas according to the views here presented it is restricted to the case of rigid dielectrics. We shall consider especially the fluid dielectric for which Helmholtz gives the force (1). We first transform (24).

Write

$$U = \frac{1}{2} \int_{v_1} \rho \phi dv. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

From (21)

$$\int_{v_2} \left\{ \mathbf{P} \cdot \frac{\partial \mathbf{E}}{\partial x_\alpha} u_\alpha + \mathbf{E} \cdot \frac{\partial}{\partial x_\alpha} (\mathbf{P} u_\alpha) \right\} dv = \int_{v_2} \frac{\partial}{\partial x_\alpha} (\mathbf{P} \cdot \mathbf{E} u_\alpha) dv = \int_{f_2} \mathbf{P} \cdot \mathbf{E}_\alpha u_\alpha df.$$

Then

$$\Delta U = \Delta U - \frac{1}{2} \int_{v_2} (\mathbf{P} \cdot \delta \mathbf{E} + \mathbf{E} \cdot \delta \mathbf{P}) dv - \frac{1}{2} \int_{f_2} \mathbf{P} \cdot \mathbf{E}_\alpha u_\alpha df.$$

As $\mathbf{P} = \mathbf{p}$ when $t=0$ we have by (10) and (17)

$$\begin{aligned} \Delta U &= \Delta U - \frac{1}{2} \int_{v_2} (\mathbf{p} \cdot \Delta \mathbf{E} + \mathbf{E} \cdot \Delta \mathbf{p}) dv + \frac{1}{2} \int_{v_2} \left\{ \mathbf{P} \cdot \frac{\partial \mathbf{E}}{\partial x_\alpha} u_\alpha + \mathbf{E} \cdot \frac{\partial}{\partial x_\alpha} (\mathbf{P} u_\alpha) \right\} dv \\ &\quad - \frac{1}{2} \int_{f_2} \mathbf{P} \cdot \mathbf{E}_\alpha u_\alpha df \\ &= \Delta U - \frac{1}{2} \int_{v_2} (\mathbf{p} \cdot \Delta \mathbf{E} + \mathbf{E} \cdot \Delta \mathbf{p}) dv. \end{aligned}$$

Hence from (24)

$$\Delta W = -\Delta U + \frac{1}{2} \int_{v_2} (\mathbf{p} \cdot \Delta \mathbf{E} - \mathbf{E} \cdot \Delta \mathbf{p}) dv. \quad . \quad . \quad . \quad . \quad (30)$$

For a dielectric held rigid in which $\mathbf{p} = \mathbf{P} = k\mathbf{E}$ we have $\mathbf{p} \cdot \Delta \mathbf{E} - \mathbf{E} \cdot \Delta \mathbf{p} = 0$. Then $\Delta W = -\Delta U$ so that U is a mechanical potential energy for the system. It is the ordinary electrostatic energy. In general when deformations of the dielectric are permitted, $\mathbf{p} = \nu \mathbf{P} = \nu k \mathbf{E}$ and $\Delta \mathbf{p} = \nu k \Delta \mathbf{E} + \mathbf{E} \Delta(\nu k)$,

$$\mathbf{p} \cdot \Delta \mathbf{E} - \mathbf{E} \cdot \Delta \mathbf{p} = -E^2 \Delta(\nu k). \quad . \quad . \quad . \quad . \quad (31)$$

If k be a function of the density τ only, we have, using (18) and taking $t=0$, $\nu=1$,

$$\begin{aligned} \Delta(\nu k) &= \nu \Delta k + k \Delta \nu \\ &= \frac{\partial k}{\partial \tau} \Delta \tau + k \Delta \nu \\ &= \left(-\tau \frac{\partial k}{\partial \tau} + k \right) \Delta \nu \\ &= -\tau^2 \frac{\partial}{\partial \tau} \left(\frac{k}{\tau} \right) \frac{\partial u_\alpha}{\partial x_\alpha}. \quad . \quad . \quad . \quad . \quad (32) \end{aligned}$$

Suppose the deformation involves the dielectric only. Expressing the work done directly in terms of the volume force \mathbf{F} and surface traction \mathbf{T} we have from (30), (31), (32)

$$\int_{v_2} \mathbf{F} \cdot \mathbf{u} dv + \int_{f_2} \mathbf{T} \cdot \mathbf{u} df = -\Delta U + \frac{1}{2} \int_{v_2} E^2 \tau^2 \frac{\partial}{\partial \tau} \left(\frac{k}{\tau} \right) \frac{\partial u_\alpha}{\partial x_\alpha} dv.$$

On Helmholtz's theory the corresponding force $\mathbf{F}^{(h)}$ and traction $\mathbf{T}^{(h)}$ are derived from the equation

$$\int_{v_1} \mathbf{F}^{(h)} \cdot \mathbf{u} dv + \int_{f_2} \mathbf{T}^{(h)} \cdot \mathbf{u} df = -\Delta U.$$

By subtraction

$$\begin{aligned} \int_{v_1} (\mathbf{F}^{(h)} - \mathbf{F}) \cdot \mathbf{u} dv + \int_{f_2} (\mathbf{T}^{(h)} - \mathbf{T}) \cdot \mathbf{u} df &= -\frac{1}{2} \int_{v_1} E^2 \tau^2 \frac{\partial}{\partial \tau} \left(\frac{k}{\tau} \right) \frac{\partial v_\alpha}{\partial x_\alpha} dv \\ &= \frac{1}{2} \int_{v_1} \nabla \left(E^2 \tau^2 \frac{\partial}{\partial \tau} \left(\frac{k}{\tau} \right) \right) \cdot \mathbf{u} dv - \frac{1}{2} \int_{f_2} E^2 \tau^2 \frac{\partial}{\partial \tau} \left(\frac{k}{\tau} \right) \mathbf{n} \cdot \mathbf{u} df. \end{aligned} \quad (33)$$

As \mathbf{u} may be arbitrary inside v_2 we infer

$$\mathbf{F}^{(h)} - \mathbf{F} = \frac{1}{2} \nabla \left(E^2 \tau^2 \frac{\partial}{\partial \tau} \left(\frac{k}{\tau} \right) \right)$$

and

$$\mathbf{T}^{(h)} - \mathbf{T} = -\frac{1}{2} E^2 \tau^2 \frac{\partial}{\partial \tau} \left(\frac{k}{\tau} \right) \mathbf{n}.$$

The first result is (4) and the second gives the surface traction according to Helmholtz. This derivation of his results differs from the usual one and we see clearly the error in his method according to our views. We remark that if in (33) we take \mathbf{u} to be an arbitrary constant vector we can infer

$$\int_{v_1} \mathbf{F}^{(h)} dv + \int_{f_2} \mathbf{T}^{(h)} df = \int_{v_1} \mathbf{F} dv + \int_{f_2} \mathbf{T} df.$$

This shows that our theory and that of Helmholtz give the same total force on the dielectric.

§ 6. THE CONSERVATION OF ENERGY.

Let us consider a fluid dielectric in equilibrium under electrical influence. Then the mechanical forces of electric origin acting in the fluid must be balanced by a hydrostatic pressure ϖ developed in the fluid: if the electric field be gradually established in the fluid the latter will be deformed so as to generate the pressure required to compensate the above mechanical forces. Thus the pressure ϖ satisfies the equation

$$\nabla \varpi = \mathbf{F},$$

where \mathbf{F} is the force (2). Let us suppose the fluid occupies the volume v_2 and that it is kept in equilibrium by a suitable normal pressure Π applied to its boundary f_2 . Then

$$(\Pi - \varpi) \mathbf{n} = \mathbf{T},$$

where \mathbf{T} is the force (20). We suppose the electric influence is supplied by the charge distribution in v_1 .

Consider an infinitesimal deformation \mathbf{u} of the dielectric, keeping the charge distribution in v_1 fixed. Let ΔQ be the heat absorbed by the dielectric in the variation and let ΔI be the increment of its internal energy.

Then ΔV is the increment of electrical energy of the system and

$$\int_{f_2} \Pi \mathbf{n} \cdot \mathbf{u} df$$

is the work done by the system against the external pressure Π . Conservation of physical energy requires that

$$\begin{aligned} \Delta Q &= \Delta I + \Delta V + \int_{f_2} \Pi \mathbf{n} \cdot \mathbf{u} df \\ &= \Delta I + \Delta V + \int_{f_2} \varpi \mathbf{n} \cdot \mathbf{u} df + \int_{f_2} \mathbf{T} \cdot \mathbf{u} df. \end{aligned}$$

Now

$$\begin{aligned} \int_{f_2} \varpi \mathbf{n} \cdot \mathbf{u} df &= \int_{f_2} \varpi n_\alpha u_\alpha df = \int_{v_2} \frac{\partial}{\partial x_\alpha} (\varpi u_\alpha) dv \\ &= \int_{v_2} \frac{\partial \varpi}{\partial x_\alpha} u_\alpha dv + \int_{v_2} \varpi \frac{\partial u_\alpha}{\partial x_\alpha} dv \\ &= \int_{v_2} \nabla \varpi \cdot \mathbf{u} dv + \int_{v_2} \varpi \Delta v dv \\ &= \int_{v_2} \mathbf{F} \cdot \mathbf{u} dv + \int_{v_2} \varpi \Delta v dv. \end{aligned}$$

Thus

$$\begin{aligned} \Delta Q &= \Delta I + \Delta V + \int_{v_2} \mathbf{F} \cdot \mathbf{u} dv + \int_{f_2} \mathbf{T} \cdot \mathbf{u} df + \int_{v_2} \varpi \Delta v dv \\ &= \Delta I + \Delta V + \Delta W + \int_{v_2} \varpi \Delta v dv, \end{aligned}$$

where ΔW is the work done by the mechanical forces of electric origin during the deformation. Hence by (24)

$$\Delta Q = \Delta I + \int_{v_2} \varpi \Delta v dv - \int_{v_2} \mathbf{E} \cdot \Delta \mathbf{p} dv.$$

We infer the elementary relation

$$\Delta q = \Delta i + \varpi \Delta v - \mathbf{E} \cdot \Delta \mathbf{p}, \quad \dots \dots \dots (34)$$

where i is the internal energy and Δq is the heat absorbed per unit volume at any place in the dielectric. Clearly Δv is a volume increment reckoned per unit volume. The equation (34) has been used as the expression of the first law in the application of thermodynamics to dielectric phenomena; here it appears as a formal derivate from the formula (2). So the formula (2) is consistent with the principle of energy in its widest sense.

In conclusion I wish to thank Professor T. G. Room, F.R.S. for his interest in and his encouragement in the writing of this paper and also Mr. E. Cunningham of St. John's College, Cambridge, for some helpful criticisms. I thank Mr. J. W. Reed for his help with the composition of an earlier version.

REFERENCES.

- ABRAHAM-BECKER, 1932, *Theory of Electricity*, 1, 91.
 HELMHOLTZ, 1881, *Weid. Ann.*, 13, 385.
 JEANS, 1927, *Theory of Electricity*, p. 172.
 LARMOR, 1897, *Phil. Trans. A*, 190, 280.
 LIVENS, 1926, *Theory of Electricity*, Chapter II.
 PIDDUCK, 1925, *Treatise on Electricity*, p. 91.
 POCKELS, 1906, *Encyklopädie der math. Wiss.*, V, Part II.

XLIV. *Notices of New Books and Periodicals received.*

The Freezing of Supercooled Water. By N. ERNEST DORSEY. [Trans. Amer. Phil. Soc., Vol. 38, part 3, pp. 82.] (American Philosophical Society, Philadelphia, November, 1948.) Price \$1.75.

THIS is a long scientific paper in which Dr. Dorsey of the U.S. National Bureau of Standards, author of "Properties of the Ordinary Water-Substance", reports *in extenso* the results of 12 years intermittent study of the supercooling of this his favourite material. Since the observations on this interesting subject are scattered and of variable reliability, it is useful to have a painstaking record of the basic facts, even though none are strikingly new.

The author demonstrates that the nucleation of ice in his samples of water (of very varied origin and purity) occurs invariably on foreign particles (at any temperature, usually characteristic of the sample for a short period, between -2.7° and -22° C.). After critical comment on previous theories (referring for example to "unwarranted reverence for physical conclusions announced by mathematicians"), he proposes his own new theory of nucleation, that the embryos of ice are portions of an adsorption layer of water molecules torn from the surface of "motes" of optimum convexity by impact of fast molecules, or from surfaces of other than optimum convexity by cavitation or rubbing. There is a valuable bibliography. F. C. F.

Science and Technology in China. Published bi-monthly by the National Science Society of China.

THE editors have received Nos. 1 and 2 of Volume I of this periodical, published in Nanking in April, 1948. These numbers contain some articles on topics of general interest, such as the work of UNESCO in China, scientific activities of the British Council and the outlook for engineering research, as well as some articles giving results of research and field work.

A History of Science, and its Relations with Philosophy and Religion. By Sir WILLIAM CECIL DAMPIER, Sc.D., F.R.S. Fourth Edition. Cambridge University Press. 25s. net.

AT a time when interest in the teaching of the history of science is growing in our universities, the appearance of a fourth edition of Sir William Dampier's famous book is particularly welcome. Much of the book is the same as the edition of 1942; but the chapter of that edition dealing with developments of the period 1930-40 has been deleted, and its subject-matter spread among earlier chapters where they more properly belong. The book also contains some account of scientific developments during the war, in so far as these have led to an increase in knowledge. But the book is eminently a treatise on pure science and not on its applications; so applied a subject as the military applications of atomic energy receives a paragraph only.

Essays in Philosophy and Other Pieces. By A. D. RITCHIE. [Pp. vi+208.] (London: Longmans, Green & Co., 1948.) Price 12s. 6d.

IN this volume Professor Ritchie has collected fifteen essays, most of them reprinted with minor changes from contributions to philosophical journals during the last ten years. In the author's classification they fall into five groups. There are "popular" expositions of subjects of general interest: freedom (of choice in action), pacifism, miracles, magic in modern politics, and the biological approach to philosophy. The essays on logical positivism, the atomic theory as metaphysics and science, the logic of question and answer, and Aristotle's logic are more technical. Theological themes are discussed in an essay on theories of immortality, and in a pseudo-Socratic dialogue on the existence of God. A tercentenary essay on Newton and a tribute to Samuel Alexander are based on public lectures. Finally, there are two short descriptive pieces, evocative cameos of the tiny island of Dochet, and of almost deserted Letang, visited during summers in Eastern Canada.

The felicity of phrasing makes all Ritchie's work a pleasure to read. In the philosophical essays the analysis of arguments in favour of diverse views is, however, sometimes presented so neutrally that it is difficult to see just what the author is driving at, but that the interest is in the discussion rather than in any conclusion is perhaps appropriate for many of the themes considered. Here and there the author does not hesitate to state his own opinion. On phenomenalism, for example, "Whatever the truth may be," he says, "this theory I am sure is false and the mother of a great family of fallacies." So wide a range of subject-matter makes general criticism and appraisal unusually difficult. It is perhaps fair to say that where the author, as often, is dealing with particular aspects of a wide theme, the general perspective is not well brought out, and it is not clear what is novel and what is simply a restatement of older views.

To the reviewer, the essays which gave most satisfaction were that on the errors of logical positivism, which gives scope for the author's clear-headed analytical powers, and that on Samuel Alexander, where affection and admiration for a man are blended harmoniously with constructive criticism of his philosophical work. There is, however, stimulus and enlightenment in all the essays, and many readers will welcome the discussions, brief though they may be, of so many vital problems by such an acute thinker and clear writer as Professor Ritchie.

E. C. S.

Journal of the Physical Society of Japan.

THE editors have received Volume III, No. 1 of this journal, issued in Tokio in January, 1948. The contents include papers on the theory of the thermionic constants of metals, applications of the collective electron theory to ferromagnetism, crystal rectifiers, properties of ultrasonic waves and the theory of the magnetron.

The Collected Works of J. Willard Gibbs, in two volumes: Volume I.—Thermodynamics; Volume II.—Elementary Principles of Statistical Mechanics. Yale University Press (London: Geoffrey Cumberlege). 45s. net.

THIS is a reprint of the first complete edition of Willard Gibbs' works, which Longley and Van Name of Yale University, where Gibbs was Professor, brought out in 1928 following an earlier incomplete edition in 1906. The contributions made by Willard Gibbs to thermodynamics are so profound that his works are in no way out of date and are essential for any serious student of the subject.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

XLV. *Magneto-thermal Effects in Ferromagnetics.*

By E. C. STONER, F.R.S., and P. RHODES,
Physics Department, University of Leeds*.

[Received October 18, 1948.]

§ 1. INTRODUCTION.

THE longest and best known of all magneto-thermal effects is the heating which occurs on taking a ferromagnetic through a magnetization cycle under adiabatic conditions, an effect essentially associated with irreversible magnetization processes. Though the aggregate rise of temperature when a single cycle is traversed adiabatically is very small, of the order of 10^{-4} deg. C. for soft magnetic materials, the hysteresis loss may be determined with considerable accuracy by a number of well-known methods. The measurement of the temperature changes, due to both reversible and irreversible processes, as a cycle, or an initial curve, is traversed step by step, is much more difficult. The first successful experiments seem to be those of Adelsberger in 1927, and since then several further investigations have been made (see § 8), the most recent being those of Bates and his collaborators (1941 onwards), which have provided detailed information about a wide range of materials. There is a remarkable diversity in the course of the temperature changes and in their magnitude, but hitherto little more than brief qualitative discussions have been given of the broad features of the curves relating temperature with field, or with magnetization (see, for example, Becker and Döring, 1939, pp. 267-70). In this paper a quantitative treatment of the basic effects is presented.

The thermodynamic relations, which are first considered (§ 2), serve to link the adiabatic temperature change characteristics of a material with other characteristics which are better known, or more readily investigated. Moreover, although these relations can give no information about the atomic details of the physical changes involved, they are invaluable in the discussion of particular types of physical process which are presumed to occur as a ferromagnetic changes its state of magnetization. A change of magnetization is, in general, the resultant effect of elementary changes of several different kinds, and the only practicable procedure seems to be to consider the various elementary processes at first separately. The information obtained about their thermal characteristics can then be applied to the interpretation of experimental temperature, field curves for different materials. There are three main processes by which change of apparent magnetization may occur: change in degree of alignment of the

* Communicated by the Authors.

elementary carriers of the magnetic moment, giving a change in the intrinsic magnetization; rotation of domain magnetization vectors; and translational motion of domain boundaries. After a brief general survey of these elementary magnetization processes (§ 3), the thermal effects associated with each of them, when taking place reversibly, are discussed (§§ 4, 5, 6). Irreversible changes and hysteresis effects are then briefly considered (§ 7). The experimental results on the adiabatic temperature changes accompanying magnetization are reviewed, and the quantitative relations obtained for the various elementary processes are utilized in developing an interpretation of the significance of the observed temperature changes for a number of representative materials (§ 8). Finally, specific ways in which measurements of the temperature changes may contribute to a fuller understanding of ferromagnetic phenomena are indicated (§ 9).

§ 2. THE THERMODYNAMICS OF MAGNETIZATION.

Certain aspects of the thermodynamics of magnetization were considered in two papers published some years ago (Stoner, 1935, 37), and the general method of formulation there adopted seems especially suitable when the "system of interest" is particular magnetizable material in an external magnetic field. (For other treatments which may be more appropriate in dealing with the general energy relations in a closed electromagnetic system, reference may be made to the papers of Guggenheim (1936), and of Livens (1945, 47, 48).) Of the various energy functions, the relations between which were discussed, it seems most convenient to take as basic an internal energy function, denoted by E , and defined by the equation

$$dE = dQ + H dI, \quad \dots \dots \dots (2.1)$$

where I is the volume intensity of magnetization, dQ the heat transferred to the body, and E and Q refer to unit volume. The field H is the external field corrected for demagnetization (the "pipe" field), and E , correspondingly, does not include the demagnetizing field energy. (The function E , accordingly, is characteristic of the material, but independent of the shape of the particular specimen on which measurements are made. Experimental results, as given, are usually corrected for demagnetization effects.) Products such as HdI are throughout scalar products of vectors, and it does not seem necessary to use a special vector notation. Strictly, (2.1) is appropriate only for changes in which the volume and shape of the specimen are fixed, but for the main applications in view the difference in the energy change at constant volume and constant (atmospheric) pressure may be neglected. If the effect of an applied stress is under consideration it is necessary to modify (2.1) by the introduction of the corresponding stress, strain term.

The heat intake, dQ , is related to the entropy increase, dS , by

$$dQ \leq T dS, \quad \dots \dots \dots (2.2)$$

the equality sign applying when the changes are reversible. Reversibility over a given range is the practical counterpart of the theoretical possibility of a unique specification of the state of the system over that range in terms

of a limited number of variables. Thus, under reversible conditions, E is a thermodynamic function appropriate to the variables S and I . From E , three other functions may be obtained, namely, $E-IH$, $E-TS$, and $E-IH-TS$, appropriate to other pairs of variables. These functions will be denoted by E' , F and F' respectively; E and E' are "internal" and "total" energy functions ("corrected" for demagnetization effects), and F and F' the corresponding free energy functions. It may be noted that F' is the function usually referred to as the free energy in statistical treatments. For reversible changes the four functions satisfy the following relations:

$$dE = TdS + HdI, \quad (2.3)$$

$$dE' = TdS - IdH, \quad (2.4)$$

$$dF = -SdT + HdI, \quad (2.5)$$

$$dF' = -SdT - IdH. \quad (2.6)$$

For irreversible changes the left-hand side in these equations is less than the right-hand side, as follows from (2.1) and (2.2) and the definitions of the functions. The reciprocity relations corresponding to (2.3) to (2.6) are

$$(\partial T / \partial I)_S = (\partial H / \partial S)_I, \quad (2.3a)$$

$$(\partial T / \partial H)_S = -(\partial I / \partial S)_H, \quad (2.4a)$$

$$(\partial S / \partial I)_T = -(\partial H / \partial T)_I, \quad (2.5a)$$

$$(\partial S / \partial H)_T = (\partial I / \partial T)_H. \quad (2.6a)$$

Writing ρ for the density, and C_I , C_H for the specific heats at constant magnetization and at constant field respectively, the following equations, required in the subsequent discussion, are readily obtained for the isentropic change of temperature with magnetization and with field:

$$\left(\frac{\partial T}{\partial I} \right)_S = \frac{T}{\rho C_I} \left(\frac{\partial H}{\partial T} \right)_I = -\frac{T}{\rho C_I} \left(\frac{\partial S}{\partial I} \right)_T; \quad (2.7)$$

$$\left(\frac{\partial T}{\partial H} \right)_S = -\frac{T}{\rho C_H} \left(\frac{\partial I}{\partial T} \right)_H = -\frac{T}{\rho C_H} \left(\frac{\partial S}{\partial H} \right)_T. \quad (2.8)$$

The relations between the changes of temperature with magnetization and with field, and of each of these with the change of internal energy with magnetization, are expressed by the equations

$$\left(\frac{\partial T}{\partial I} \right)_S = \frac{C_H}{C_I} \left(\frac{\partial H}{\partial I} \right)_T \left(\frac{\partial T}{\partial H} \right)_S = -\frac{1}{\rho C_I} \left\{ \left(\frac{\partial E}{\partial I} \right)_T - H \right\}. \quad (2.9)$$

The specific heat at constant magnetization, C_I , and that at constant field, C_H , differ increasingly from each other as the Curie temperature, θ , is approached. At temperatures well below θ , however, as in most of the applications to be made, the relative difference is very small, and C_I and C_H may both be taken, with negligible error, as equal to the specific heat, C , as ordinarily measured in zero field.

The observational results on the thermal change accompanying adiabatic magnetization are often expressed not in terms of the temperature change but, as a matter of convenience in relation to the calibration procedure adopted, in terms of the corresponding heat or energy change (expressed, for example, in erg cm.⁻³). Denoting what may be called the heat "developed" corresponding to a temperature change ΔT by $\Delta Q'$,

$$\Delta Q' = C\rho \Delta T. \quad (2.10)$$

The equations for $(\partial Q'/\partial I)_S$ and $(\partial Q'/\partial H)_S$ corresponding to (2.7) and (2.8), can at once be written down, and are formally somewhat simpler.

Under adiabatic conditions the basic equation, (2.1), becomes

$$dE = H dI, \quad (2.11)$$

which holds whether the change of magnetization with field is reversible or irreversible. The reciprocity relations obtained above apply only if the change is reversible, that is, isentropic as well as adiabatic. For irreversible changes it is necessary to use the inequality sign in (2.2), and no obviously useful relations follow from the manipulation of the inequalities which take the place of (2.3) to (2.6). Whether or not the change is reversible, however, the change of energy may conveniently be divided into two parts, one associated with the change of temperature and the other with the change of magnetization:

$$dE = (\partial E/\partial T)_I dT + (\partial E/\partial I)_T dI. \quad (2.12)$$

This gives, with (2.11) and (2.10),

$$(\partial E/\partial I)_T dI = H dI - dQ', \quad (2.13)$$

an equation equivalent to that given by Bates and Weston (1941), and extensively used in connection with the presentation of the experimental results. Using the symbol ΔE_M for the (finite) change in energy associated with the change in I (the subscript connoting the "magnetic" as distinguished from the thermal energy change), (2.13) may be re-expressed in the integral form

$$\Delta E_M = \int (\partial E/\partial I)_T dI = \int H dI - \Delta Q', \quad (2.14)$$

and it is to be noted that ΔE_M is unambiguously determinable from the experimental data. Thus the total change in internal energy, ΔE , and the magnetic and thermal parts, ΔE_M and ΔE_T , are all determinable; in contrast the change in entropy cannot in general be determined if the process is irreversible, unless an alternative reversible method of passing from the initial to the final state can be devised. If (2.13) is written in the form

$$\left(\frac{\partial Q'}{\partial I}\right)_Q = \rho C \left(\frac{\partial T}{\partial I}\right)_Q = - \left\{ \left(\frac{\partial E}{\partial I}\right)_T - H \right\}, \quad (2.15)$$

the close similarity with one of the equations of (2.9) becomes apparent. In a sense (2.15), which holds for any actual change, subject to $dQ=0$, includes (2.9) as a special case when $dQ=T dS$, and the change is reversible.

The difficulties in connection with ferromagnetics do not, however, arise so much from the occurrence of irreversible processes as from the internal complexity of the material, which makes it impossible, in general, to specify an equilibrium (or quasi-equilibrium) state uniquely by a small number of parameters.

§ 3. ELEMENTARY MAGNETIZATION PROCESSES.

A ferromagnetic may be regarded as made up of domains, 1, 2, j ,, of volumes $v_1, v_2, \dots v_j, \dots$, in each of which, at temperatures well below the Curie point, θ , the absolute value of the intensity of magnetization approximates closely to the quasi-saturation magnetization. The intrinsic magnetization, or micro-magnetization, which will be denoted by I_m , varies with temperature and field, but at temperatures not too near θ (e.g. for iron, nickel and cobalt, at room temperature) the variation of I_m with H in fields up to several thousand oersted is very small. (For nickel, at room temperature, the difference between I_m for $H=0$, that is the spontaneous magnetization, I_0 , and I_m for $H=10^4$ is probably less than 0.3 per cent.) None the less, the associated magneto-thermal effects may account for a large part of the observed adiabatic temperature change. The mean magnetic moment per unit volume, I , corresponding to the holomagnetization as ordinarily measured, may be written

$$I = \sum_j v_j (I_m)_j \cos \phi_j, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.1)$$

where ϕ is the angle between I_m and I , and the summation is over the domains in unit volume. Change of I may therefore be expressed as a resultant effect of changes in I_m , ϕ , and v :

$$dI = \sum_j \{ v_j \cos \phi_j d(I_m)_j - v_j (I_m)_j \sin \phi_j d\phi_j + (I_m)_j \cos \phi_j dv_j \}. \quad . \quad (3.2)$$

The three types of change correspond to change in intrinsic magnetization, rotation of the domain magnetization vector, and domain boundary movements, to be considered in subsequent sections. The heat developed, dQ' per unit volume, is a resultant of the contributions from the three types of elementary process, and may be formally expressed by

$$dQ' = \sum_j \left\{ \frac{\partial Q'}{\partial (I_m)_j} d(I_m)_j + \frac{\partial Q'}{\partial \phi_j} d\phi_j + \frac{\partial Q'}{\partial v_j} dv_j \right\}. \quad . \quad . \quad . \quad (3.3)$$

The above equations will make clear the ideal requirements for a quantitative interpretation of the observed adiabatic temperature changes accompanying changes in magnetization. They are firstly the determination, from other data or from theory, of the heat developed in the various elementary processes, and secondly the estimate of the contribution from the various processes to the observed change in magnetization. Unfortunately, such estimates can in general be made only very roughly, and, with regard to the first requirement, there are often uncertainties as to the change of magnetization with field for a particular type of process.

It may then be possible to estimate with fair precision from available data the change of temperature with change of field, but not with change of magnetization. Owing to these complications it is not practicable to complete the treatment of the thermal effects associated with change in intrinsic magnetization in §4 before considering rotational and translational processes in §§5 and 6, and, similarly, rotational effects are considered in §6 as well as in §5. The three sections deal with reversible changes only, and irreversible processes are considered together in §7.

§4. CHANGE IN INTRINSIC MAGNETIZATION.

Magneto-caloric effect in high fields.

The adiabatic change in temperature accompanying magnetization in high fields is not distinct in character from that in low and moderate fields except in so far as it may be associated almost entirely with change in intrinsic magnetization. Detailed measurements have been made of this magneto-caloric effect over a wide range of temperatures in nickel by Weiss and Forrer (1926) and in iron by Potter (1934). For the same change of field the temperature change increases as the Curie point is approached; in the Curie point region, for field changes of about 10^4 oersted, temperature changes of more than 1° are observed in nickel and of more than 2° in iron. Temperature changes were measured by methods sensitive to about 10^{-3} degree, a sensitivity adequate for the purpose of the particular investigations, but insufficient for the measurement of the small changes in lower fields at room temperature. It is, however, necessary to indicate some of the conclusions reached in a consideration of the high field results, as they have an important bearing on the interpretation of the thermal changes in low fields.

The rate of change of temperature with apparent magnetization, $(\partial T/\partial I)_S$, is small until I approaches the "technical saturation" value, after which it increases rapidly. Ultimately, the aggregate temperature change, ΔT , increases approximately linearly with the square of the apparent magnetization, being then given by

$$\Delta T = a(I^2 - b), \quad \dots \dots \dots (4.1)$$

where b , a constant for each particular temperature, may plausibly be identified with the square of the spontaneous magnetization, I_0^2 , at that temperature, and a would be expected to be approximately proportional to the Weiss molecular field coefficient. Potter's analysis of his results suggested that the molecular field coefficient decreased fairly rapidly with decreasing temperature. The low values indicated at the lower temperatures are, however, incompatible with other characteristics of ferromagnetics. Partly with the aim of tracing the origin of the apparent inconsistencies, various methods of determining the dependence on intrinsic magnetization of the internal energy of ferromagnetics from experimental data (on the magneto-caloric effect, the specific heat, and the magnetic isothermals) were discussed in detail by Stoner (1936), with particular reference to nickel. It was shown that the major discrepancies

arose from the assumption that the whole measured change in magnetization in the higher field ranges could be identified with the change in intrinsic magnetization; and other processes which could contribute to the magnetization change even in high fields were indicated. The main conclusions relevant to the general interpretations of magneto-thermal effects at temperatures well below the Curie point are, firstly, that the value of $(\partial I_m / \partial H)$ may differ greatly from the experimental value obtained for $(\partial I / \partial H)$ even in high fields, even though the difference between I and I_m may be very small compared with either of them; and, secondly, that the value of $(\partial I / \partial T)_H$ in high fields gives a very close approximation to $(\partial I_0 / \partial T)$ and consequently also to $(\partial I_m / \partial T)_H$, which can vary little over a wide field range. (Reference must be made to the original paper for the full evidence for these conclusions, and for the detailed analysis of the experimental data.) The values of $(\partial I_0 / \partial T)$ can be determined with fair accuracy, and, therefore, using (2.8), reliable values of $(\partial T / \partial H)_S$ associated with the change in intrinsic magnetization can be derived; in contrast, as will be apparent from (2.9), reliable values cannot be derived for the rate of change of temperature with change in intrinsic magnetization. This is one reason why magneto-caloric effects generally cannot be dealt with in direct accordance with the ideal scheme suggested by (3.3).

Data for ferromagnetic elements.

For a quantitative treatment of the magneto-caloric effects in any particular material, various data for that material are required, data which are not always available for the materials on which the magneto-caloric measurements have been made. For purposes of illustration and

TABLE I.

Data for ferromagnetic elements at room temperature ($T = 290^\circ \text{K.}$).

θ , Curie temperature.

σ_{00} , σ_{0T} , intrinsic magnetization per gram, extrapolated to zero field, at 0°K. , $T^\circ \text{K.}$

ρ , density.

I_0 , intrinsic magnetization per cm.^3 , extrapolated to zero field, at $T^\circ \text{K.}$

C , specific heat at $T^\circ \text{K.}$, $\text{cal. deg.}^{-1} \text{gm.}^{-1}$.

$\Delta Q'$, energy change, erg cm.^{-3} , corresponding to temperature change ΔT .

	T/θ	σ_{0T}/σ_{00}	ρ	$I_0 \times 10^{-3}$	$-(dI_0/dT)$	C	$(\Delta Q'/\Delta T) \times 10^{-7}$	$(\Delta T/\Delta Q') \times 10^8$
Fe	0.278	0.982	7.9	1.72	0.20	0.107	3.5 ₃	2.8 ₃
Ni	0.460	0.944	8.9	0.48 ₅	0.24	0.105	3.9 ₁	2.5 ₆
Co	0.208	0.994	8.8	1.42	0.04 ₃	0.104	3.8 ₂	2.6 ₂

general discussion, the ferromagnetic elements iron, nickel and cobalt will be considered. The adopted values, used here in numerical work, for the specific intensities of magnetization at absolute zero, σ_{00} , are 221.7, 57.6 and 162 respectively, and for the Curie temperatures 1043, 631 and 1393°K. Data for room temperature (taken, for definiteness, as 290°K.) are given in Table I.

Values given for $(\sigma_{0T}/\sigma_{00})$ and (dI_0/dT) are based on the results of Weiss and Forrer (1926, 29), except (dI_0/dT) for cobalt, which is derived from data kindly supplied by Professor W. Sucksmith and Mr. H. P. Myers, Sheffield University. Their results, although not final, are undoubtedly much more reliable than any previously available. The values adopted for the density, ρ , and specific heat, C , are taken from standard tables.

Adiabatic temperature changes.

Using the data in Table I., the adiabatic change in temperature with field $(\partial T/\partial H_p)_s$, associated with reversible increase in intrinsic magnetization may at once be obtained from (2.8), and also the corresponding heat developed, $(\partial Q'/\partial H_p)_s$, using (2.10), with the results shown in Table II. The subscript p is used to denote that H is parallel to I_m , which is implicit in the method of derivation.

TABLE II.

Estimated adiabatic temperature change at room temperature associated with change in intrinsic magnetization.

Q' , heat developed, erg cm.⁻³.

$(\partial T/\partial H_p)_s$, deg. oersted⁻¹.

$(\partial Q'/\partial H_p)_s$, erg cm.⁻³ oersted⁻¹.

	$(\partial T/\partial H_p)_s \times 10^6$	$(\partial Q'/\partial H_p)_s$
Fe	1.6 ₄	58
Ni	1.7 ₈	70
Co	0.3 ₃	12

The estimated values of $(\partial T/\partial H_p)$ and $(\partial Q'/\partial H_p)$ in Table II., obviously approximate to the values of $(\partial T/\partial H)$ and $(\partial Q'/\partial H)$ for a specimen as a whole associated with change in I_m only when there is approximate parallelism of all the domain magnetization vectors and the field, *i.e.* in the saturation region. In general, using the symbolism of § 3, for changes of I_m without changes of ϕ_j or v_j ,

$$\left(\frac{\partial Q'}{\partial H}\right)_s = \sum v_j \left(\frac{\partial Q'}{\partial H_p}\right) \left(\frac{\partial H_p}{\partial H}\right)_j = \left(\frac{\partial Q'}{\partial H_p}\right) \sum v_j \cos \phi_j = \frac{I}{I_0} \left(\frac{\partial Q'}{\partial H_p}\right)_s, \quad (4.2)$$

remembering that, for the fields and temperatures primarily to be considered, the relative differences between I_0 and I_m (and also the technical saturation value of I) are negligible. It is to be noted that (4.2) does not include the whole of the thermal effect associated with change in intrinsic magnetization. For a single domain I_m may be treated as a function of H , ϕ , and v , so that

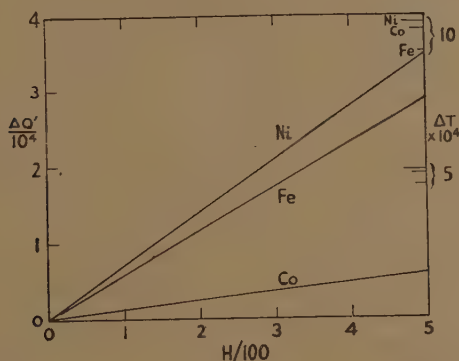
$$dI_m = \left(\frac{\partial I_m}{\partial H}\right) dH + \left(\frac{\partial I_m}{\partial \phi}\right) d\phi + \left(\frac{\partial I_m}{\partial v}\right) dv. \quad (4.3)$$

The thermal effect covered by (4.2) corresponds to the first term. The effects corresponding to the second and third terms are more conveniently considered in the sections on rotational and translational processes (§§ 5, 6).

For ease in comparison with later curves the results embodied in Table II. are shown graphically in fig. 1. From the above discussion the slope of the lines gives the estimated limiting slope when translational and rotational processes are completed. As indicated by (4.2) the slope would, in general, initially be less than that shown, and the straight lines therefore give the estimated maximum changes in Q' and T associated with reversible change in intrinsic magnetization.

The values of $(\partial T/\partial H_p)_s$ in Table II. may be taken as typical for ferromagnetic materials at temperatures well below the Curie point (see Table I. for values of T/θ). Contrary to what is sometimes supposed, the heat developed through changes in intrinsic magnetization may be a major contribution to the total observed adiabatic temperature change in

Fig. 1.



Estimated heat developed and temperature change with adiabatic change in intrinsic magnetization.

$\Delta Q'$, heat developed, erg cm⁻³.

low and moderate as well as in high fields. At higher temperatures for the ferromagnetic elements, or at room temperature for materials with lower Curie temperatures, the values of (dI_0/dT) , and consequently of $(\partial T/\partial H)_s$, may be very much greater. The relatively large magneto-caloric changes observed with certain alloys arise in this way (see § 8).

Atomic theory.

A quantitative "explanation", as distinct from a thermodynamic coordination, of the observed magneto-caloric effects associated with changes of intrinsic magnetization, may be said to have been obtained when values derived from a satisfactorily based atomic theory of ferromagnetism are in reasonable agreement with experiment. Comparison may appropriately be made between the theoretical and experimental values of (dI_0/dT) , as being more directly obtained from theory, and with greater reliability from experiment than the related magneto-caloric quantities. The simple molecular field treatment of ferromagnetism,

using Maxwell-Boltzmann statistics with the quantum modification appropriate to electron spins as elementary carriers of the magnetic moment, leads to a curve for the variation with temperature of the spontaneous magnetization below the Curie point which is of the same general form as those obtained experimentally. There are, however, serious discrepancies in the values of (dI_0/dT) at particular temperatures. For iron, for example, the theoretical value to be compared with the experimental value of 0.20 (Table I.) is 0.066. This treatment may be regarded as a first approximation to more rigorous treatments on either a Heitler-London basis, or a collective electron basis, using Fermi-Dirac statistics. Though treatments of these two types have contributed enormously, in different ways, to an understanding of ferromagnetism, neither, as so far developed, has led to a better agreement with experiment in the particular respect under consideration than the relatively crude first approximation. The fundamental theoretical treatments of ferromagnetism are, in short, insufficiently developed to be usefully applicable in the quantitative treatment of magneto-caloric effects, and it would be out of place to consider them in detail here. It is the less necessary in that these various treatments have recently been fairly fully reviewed (Stoner, 1948). In spite of the present quantitative deficiencies of the atomic theory of ferromagnetism, a statistical treatment, even in its simplest form, in which both entropy and magnetization are interpreted in terms of the distribution of "particles" among available states, is of central importance to the understanding of all reversible magneto-caloric effects. It is, therefore, proper to indicate how the relations between changes in entropy and magnetization are interpreted in terms of atomic theory.

Entropy changes.

The isentropic changes of temperature with magnetization and with field are related to the isothermal changes of entropy by (2.7) and (2.8). Under isothermal conditions the entropy change is that associated with the magnetic part of the system, due to a change in the distribution of the elementary carriers of the magnetic moment; under isentropic conditions this change is balanced by a change in the entropy associated with the lattice vibrations. A brief indication will be given of the character of the atomic processes underlying the thermodynamic relations.

It is sufficient to consider a paramagnetic in which the carriers of the magnetic moment are effectively electron spins ($J=\frac{1}{2}$, $g=2$, magnetic moment $\mu=\mu_B$). The relative magnetization, ζ , is given by

$$\zeta = I/I_0 = \tanh(\mu H/kT) = \tanh \beta. \quad . \quad . \quad . \quad (4.4).$$

An expression for the difference between the entropy in a field H , S , and in zero field, S_0 , may be obtained in the form

$$(S-S_0)/nk = \phi = \ln \cosh \beta - \beta \tanh \beta, \quad . \quad . \quad . \quad (4.5).$$

where n is the effective number of spins in the volume to which S refers..

For $\beta \rightarrow 0$,

$$\phi \rightarrow -\frac{1}{2}\beta^2 + \frac{1}{4}\beta^4 \dots, \quad (4.5a)$$

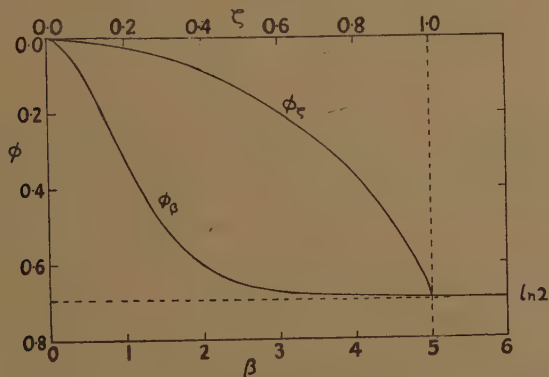
and for $\beta \rightarrow \infty$ (corresponding to $\zeta \rightarrow 1$),

$$\phi \rightarrow -\ln 2 + (1-2\beta)e^{-2\beta} \dots \quad (4.5b)$$

The variation of entropy with field and magnetization is shown graphically in fig. 2.

In zero field the spins are in a doubly degenerate state. The degeneracy is removed by a field, and with increasing field the separation increases between the two non-degenerate states, corresponding to spins parallel and anti-parallel with the field. Equality of the number of spins in the two states is incompatible with temperature equilibrium with the atomic lattice. The spin entropy decreases with increasing number of spins in

Fig. 2.



Entropy, magnetization and entropy, field curves for a paramagnetic with independent electron spins. Maxwell-Boltzmann statistics.

$$\zeta = I/I_0 = \tanh \beta.$$

$$\beta = \mu H/kT.$$

$$\phi = (S - S_0)/nk.$$

ϕ_ζ , ϕ_β , reduced entropy plotted against β , ζ .

one of the two states. The ultimate change in the reduced entropy, ϕ , of $-\ln 2$, corresponds to the change from a distribution with half the spins in each of the two states to one with all the spins in one state (the lower energy state with the spins parallel to the field). Under conditions of constant entropy for the system as a whole, the energy or "heat" lost in electron transitions from the upper to the lower energy state is transferred to the lattice vibrations. The spin entropy decrease is balanced by the lattice entropy increase, the whole process manifesting itself in the increase of temperature accompanying adiabatic magnetization. It is possible, along these lines, to give a microscopic, or "atomic", interpretation of all the terms in the various thermodynamic equations relating to the magneto-caloric effect.

For a metallic ferromagnetic the curves relating entropy with intrinsic magnetization or with effective field would, with the simple molecular field treatment, be the same as those shown in fig. 2. At temperatures well below the Curie point, however, the effective field is very large even in the absence of an applied field owing to the molecular field effect arising from exchange interaction. The zero (applied) field values of ζ then approach unity (*e.g.* the values of σ_{0T}/σ_{00} in Table I.), and it is immediately apparent from the figure that though the rate of change of spin entropy (and hence, under isentropic conditions, of temperature) with magnetization may be large, the change with field will be small. The curves shown will be quantitatively changed by more elaborate statistical treatments (*e.g.* in the collective electron treatment using Fermi-Dirac statistics); but it is unlikely that the general forms of the curves would be greatly modified. The agreement or disagreement with experiment, difficult to determine directly, is likely to be similar to that indicated for the magnetization, temperature curves.

§ 5. ROTATIONAL PROCESSES.

Change in magnetization due to rotation of the domain magnetization vectors may occur to some extent over any part of the magnetization curve, but it is usually predominant over the range between the knee of the curve and technical saturation. In this section attention will be directed mainly to the thermodynamic treatment of magneto-caloric effects in polycrystalline material associated with rotations in the anisotropic crystalline field. The modifications introduced by internal strains, which, for various reasons, it is impracticable to treat in a satisfactory quantitative manner, are indicated, and finally some tentative suggestions are made about the atomic interpretation of the various effects.

Crystal anisotropy.

The magnetization curves of single crystals may be satisfactorily co-ordinated in the rotational region by writing a general expression for the internal free energy, F (see 2.5), in a form consistent with the crystal symmetry, and minimizing the total free energy, F' , equal to $F - IH$, with respect to change of direction of the magnetization vector, I_0 , in a given field. For a cubic crystal the direction dependent part of F arising from the crystal anisotropy may be written

$$F_C = K_4(\alpha_2^2\alpha_3^2 + \alpha_3^2\alpha_1^2 + \alpha_1^2\alpha_2^2) + K_6\alpha_1^2\alpha_2^2\alpha_3^2 + \dots, \quad (5.1)$$

where the α_i are the direction cosines of I_0 with respect to the cubic axes. For nickel satisfactory agreement with experiment is obtained with the K_4 term alone; for iron agreement is sometimes improved by inclusion of the second term, but the values derived for K_6 are very inconsistent (differing even in sign), and since the mean of listed values is little more than a quarter of K_4 , for the present purpose the second term may be neglected

for iron as well as for nickel. For a hexagonal crystal, isotropic in the basal plane, an appropriate expression for F_c is

$$F_c = K'_2 \sin^2 \psi + K'_4 \sin^4 \psi + \dots, \quad (5.2)$$

where ψ is the angle between I_0 and the hexagonal axis. This expression is applicable to hexagonal cobalt for which at room temperature K'_4 is about a quarter of K'_2 , so for the present purpose the first term alone is again sufficient. The values adopted here in numerical work for the leading magneto-crystalline anisotropy coefficients and their temperature variations are shown in Table III. In arriving at these most probable values numerous original papers have been consulted, but it is unnecessary to list them here, as full references are given in the detailed surveys of Becker and Döring (1939) and of Bitter (1937). The probable errors of the K_4 values for iron and nickel are about 2 and 4 per cent respectively. For the other values listed the number of independent determinations is too small to allow of an estimate of the probable error being made.

TABLE III.

Magneto-crystalline data for ferromagnetic elements at room temperature (290° K).

K_4 , K'_2 , leading anisotropy coefficients for cubic, hexagonal crystals.
See equations (5.1), (5.2). Values in erg cm.⁻³.

dK_4/dT , dK'_2/dT . Values in erg cm.⁻³ deg.⁻¹.

	K_4	dK_4/dT	$(dK_4/dT)/K_4$
Fe	4.3×10^5	-8.6×10^2	$-2.0_0 \times 10^{-3}$
Ni	-4.9×10^4	$+7.6 \times 10^2$	$-1.5_6 \times 10^{-2}$
	K'_2	dK'_2/dT	$(dK'_2/dT)/K'_2$
Co	4.1×10^6	-1.8×10^4	-4.4×10^{-3}

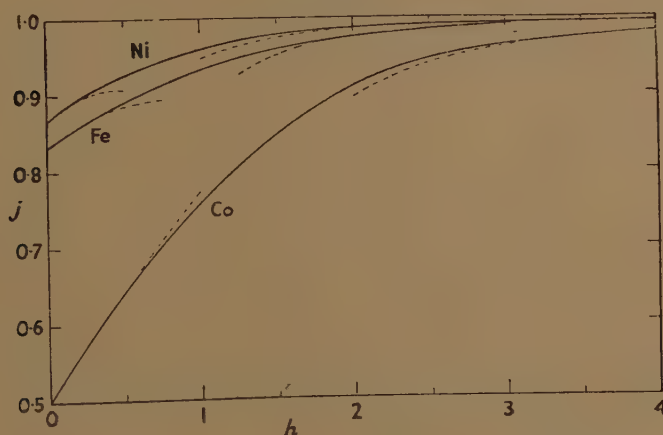
The easy directions of magnetization in a crystal, corresponding to minima of the free energy functions (5.1) and (5.2), are the six cube edge directions for iron (K_4 positive), the eight cube diagonal directions for nickel (K_4 negative), and the two hexagonal axis directions for cobalt (K'_2 positive).

Polycrystalline material.

Given the anisotropy coefficients of single crystals it is possible in principle, though it may be very troublesome in practice, to calculate the magnetization, field relations in the rotational region for an assembly of undistorted crystals with a known orientation distribution. Such calculations, for random orientation, have been made by Gans (1932) for iron, nickel and cobalt. For the purpose of the calculations it is assumed that in a vanishingly small field the magnetization vector in each crystal lies along the easy direction making the smallest angle with the field, and that as the field increases change of magnetization occurs solely by

curves are based on values obtained by a combination of graphical and numerical interpolation. As a control the values of the integral of $h dj$ from remanence to saturation obtained by numerical integration have been compared with correspondingly reduced values for the change in internal free energy obtained analytically (1/5 for iron, 2/15 for nickel). The final tables of estimated numerical values of j as a function of h from which the curves are drawn give for the integral a value agreeing with the exact value to within 1 per cent.

Fig. 3.



Estimated magnetization curves for polycrystalline material with random orientation.

$$h = HI_0/k, \quad j = I/I_0.$$

For iron, nickel, cobalt, $H/h = 250, 100, 2890$;

$$I/j = 1720, 485, 1420.$$

The broken curves give values obtained from the series derived by Gans (1932).

Adiabatic temperature changes.

Expressions can now be obtained for the adiabatic temperature change, or for the heat developed, in reversible rotational magnetization processes, in a form appropriate to the curves of fig. 3 but of more general application. The reduced magnetization, j , is to be treated as a known function of h . The variation of the intrinsic magnetization, I_m , with H , is taken into account through the dependence of I_0 on T . From (5.4) and (5.5),

$$\left(\frac{\partial I}{\partial T}\right)_H = j \frac{\partial I_0}{\partial T} + I_0 \frac{\partial j}{\partial h} \frac{\partial h}{\partial T} = j \frac{\partial I_0}{\partial T} + h \frac{\partial j}{\partial h} \frac{\partial I_0}{\partial T} - h \frac{\partial j}{\partial h} \frac{I_0}{k} \frac{\partial k}{\partial T}, \quad (5.8)$$

giving, from (2.8) and (2.10),

$$\left(\frac{\partial Q}{\partial H}\right)_s = -jT \frac{\partial I_0}{\partial T} - h \frac{\partial j}{\partial h} T \frac{\partial I_0}{\partial T} + h \frac{\partial j}{\partial h} \frac{I_0 T}{k} \frac{\partial k}{\partial T}. \quad (5.9)$$

The first term, which is positive, corresponds to increase of intrinsic magnetization without change of direction, and has already been discussed (*cf.* (4.2)); the second, also positive, gives the thermal effect of change of I_m with direction (*cf.* the second term in (4.3)); the third term, which is negative, gives the change associated purely with rotation of the magnetization vectors in the magneto-crystalline anisotropic field. It is convenient to re-express (5.9) in a form suitable for a finite change of field from, say, H_1 (or h_1) to H_2 (or h_2), associated with a change in magnetization from I_1 (or j_1) to I_2 (or j_2). In terms of the reduced variables,

$$\Delta Q' = \alpha \int j dh + \alpha \int h dj + \beta \int h dj = \alpha \int d(jh) + \beta \int h dj. \quad (5.10)$$

where

$$\alpha = -\frac{Tk}{I_0} \frac{dI_0}{dT}, \quad \beta = T \frac{dk}{dT}.$$

In terms of I and H ,

$$\Delta Q' = a \int d(IH) + b \int H dI, \quad (5.11)$$

where

$$a = \frac{\alpha}{k} = -\frac{T}{I_0} \frac{dI_0}{dT}, \quad b = \frac{\beta}{k} = \frac{T}{k} \frac{dk}{dT}.$$

Owing to the form in which results on the magneto-caloric effect are sometimes expressed, it may be useful to give explicitly the expression for the change in the magnetic part of the internal energy (*cf.* (2.14)), though it is essentially equivalent to (5.11):

$$\Delta E_M = \int H dI - \Delta Q' = -a \int d(IH) + (1-b) \int H dI. \quad (5.11a)$$

The values of the coefficients α , β , a , b , obtained from the data of Tables I. and III., are given in Table V., together with the ratios of I and H to the reduced variables j and h .

TABLE V.

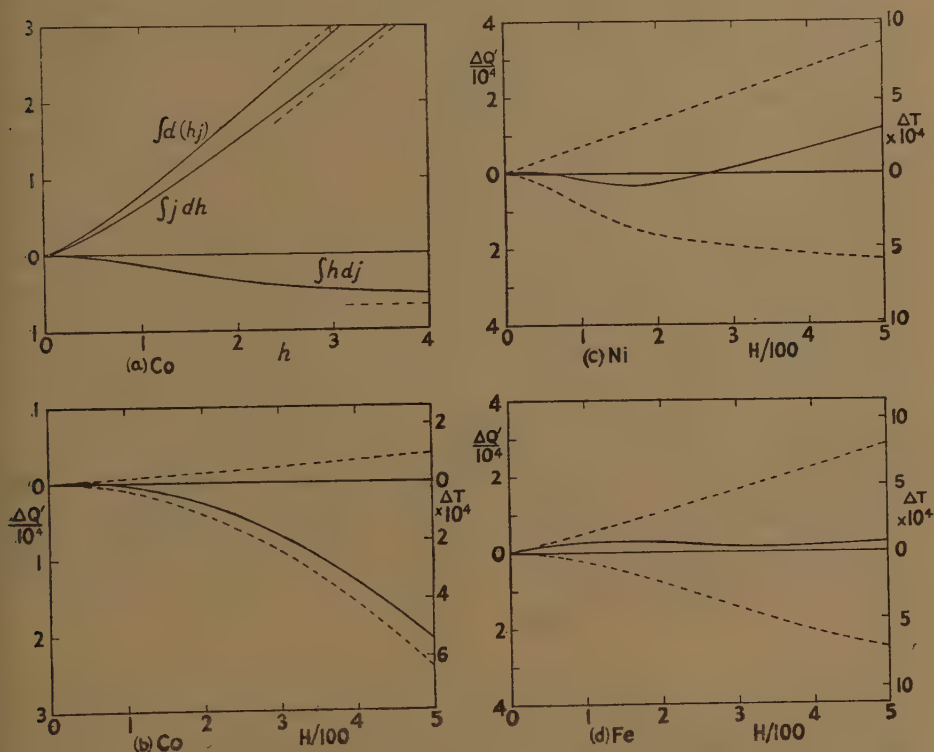
Coefficients in expressions for the adiabatic thermal changes associated with magnetization changes in unstrained polycrystalline material at room temperature (290° K.), and conversion factors. See equations (5.10), (5.11), and (5.4), (5.5).

	$\alpha \times 10^{-4}$	$-\beta \times 10^{-6}$	$a \times 10^2$	$-b$	I/j	H/h
Fe	1.45	0.25	3.4	0.58	1720	250
Ni	0.70	0.22	14.3	4.5 ₀	485	100
Co	3.6	5.2	0.8 ₈	1.2 ₇	1420	2890

The curves of fig. 4(a) give the variation with reduced field, h , of the integrals of (5.10) for cobalt, corresponding to the magnetization curves of fig. 3. The estimated curves for the variation with H of $\Delta Q'$ and ΔT are shown in fig. 4(b), and the curves for nickel and iron in figs. 4(c) and (4d).

The equation (5.11) is of central importance in the analysis of experimental results on the magneto-caloric effect, and for this reason it is appropriate to draw attention to its range of validity here, although this involves some anticipation of later sections of the paper. The first term is believed to be a valid general expression for the heat developed owing

Fig. 4.



Reduced energy integral curves for cobalt, and estimated isentropic heat and temperature changes in iron, nickel and cobalt.

The curves correspond to the magnetization curves of fig. 3.

- (a). Energy integral curves for Co (see 5.10). The broken lines are the limits of the curves. Note that $H=500$ in (b) corresponds to $h=0.17$ in (a)
- (b), (c), (d). Estimated heat ($\Delta Q'$ in erg cm. $^{-3}$) and temperature variation with H for Co, Ni and Fe. The full curves give the total effect, the upper and lower broken curves the effects associated with change in intrinsic magnetization, and with rotation of the magnetization vector in the anisotropic magneto-crystalline field.

to changes in intrinsic magnetization, whether or not the change in apparent magnetization is reversible. The accuracy of the values of the coefficients a , such as those in Table V., is, of course, dependent on the reliability of the experimental results from which I_0 and dI_0/dT are obtained; but there is no unusual experimental difficulty in determining the relevant

quantities for any particular material with an accuracy quite adequate for this application. The second term is generally valid in the rotational region for an aggregate of effectively unstrained crystalline grains, independently of their orientational distribution. The experimental determination of k and dk/dT is troublesome, and there are few materials for which data are available from which estimates of the coefficient b can be made; the estimates which can be made, such as those in Table V., are usually much less reliable than those of a . The effective anisotropy coefficients are modified by stress (see below), and the resultant modification of b usually cannot be estimated with any precision from data at present available. Moreover, in polycrystalline material, the internal stress will in general vary from grain to grain; the aggregate is then a mixture of crystals with different effective anisotropy coefficients, and the appropriate value of b may vary along the magnetization curve. It is only for ideal materials, in which the single crystals are all either unstressed or similarly stressed, that constancy of b is to be expected even over the reversible rotational region. Over regions in which irreversible changes occur, the second term in (5.11) is no longer valid even as an approximate representation of the heat developed in the reversible rotational changes which may also occur. This is because the increment in the magnetocrystalline energy no longer corresponds to the increment in the integral of HdI with a multiplying factor unity; the appropriate factor cannot be determined *a priori*, and, depending on the detailed character of the irreversible (spontaneous) changes, it may be greater or less than unity, and positive or negative. The invalidity of the second term over regions in which irreversible changes are occurring is to be contrasted with the general validity of the first term.

Strain effects.

When the magnetostriction is isotropic, the effect of a tension, Z , is to introduce into the expression for the internal free energy of a single crystal a further term of the form

$$F_s = (3/2)\lambda Z \sin^2 \phi,$$

where ϕ is the angle between I_0 and the direction of tension, and λ is the saturation magnetostriction coefficient (*cf.* Becker and Döring 1939, p. 128 ff). If λ and Z are of the same sign the qualitative effect is to increase the ease of magnetization along the direction of the stress (which eventually becomes an easy direction), and if they are of opposite sign, perpendicular to this direction. The stress effect becomes comparable with the natural crystal effect in cubic crystals (*cf.* 5.1) for

$$(3/2)|\lambda Z| = (1/3)|K_4|, \quad (5.12)$$

and in hexagonal crystals (*cf.* 5.2) for

$$(3/2)|\lambda Z| = |K'_2|. \quad (5.13)$$

Owing to the anisotropy of magnetostriction in crystals, little better than orders of magnitude can be derived from the above equations. Taking for $|\lambda|$ the maximum observed values in single crystals of about 1.8, 5.2 and 2×10^{-5} for iron, nickel and cobalt, the values of $|Z|$ satisfying (5.12) or (5.13) are about 54, 2.1 and over 1000 kg. mm.⁻² respectively. The breaking stress for external loads is of the order 200 kg. mm.⁻². The local internal stresses have been estimated as being of the order 10 kg. mm.⁻² in cold-worked polycrystalline material; in ordinarily annealed material they are undoubtedly much less. It would be expected, therefore, that the effect of internal stresses on the rotational part of the magnetization curves as estimated for completely undistorted polycrystalline material (fig. 3) would be appreciable in iron, considerable in nickel and almost negligible in cobalt.

It is virtually impossible to give a satisfactory quantitative treatment of the effect of internal stresses of arbitrary magnitude and distribution. The only general statement which seems possible is that the local stresses modify in varying degrees the effective symmetry characteristics of the crystal grains in polycrystalline material, and the effective magnitude of the anisotropy coefficients, to an extent which, in favourable cases, might be assessed by a detailed analysis of the magnetization curves.

As to the magneto-caloric effects, an equation of the form (5.11) will still hold, with the coefficient a unchanged, but with b modified, owing to the change both in the effective value of k and of its temperature coefficient. The magnitude of the modifications may be indicated by writing, as approximate expressions for the effective anisotropy and magneto-caloric coefficients,

$$k_{\text{eff}} = k + qZ\lambda = k(1 + qZ\lambda/k), \quad . \quad . \quad . \quad . \quad . \quad (5.14)$$

$$b_{\text{eff}} = b \frac{k}{k_{\text{eff}}} \left(1 + qZ \frac{d\lambda/dT}{dk/dT} \right), \quad . \quad . \quad . \quad . \quad . \quad (5.15)$$

where q is a numerical coefficient (which would ideally be equal to 9/2 and 3/2 for cubic and hexagonal crystals if a simple additivity relation held between the natural and the strain effects) and the moduli of Z and λ are to be taken. The temperature coefficient of Z may reasonably be assumed to be very small at room temperature. There are few data for $d\lambda/dT$. For nickel, however, there are some measurements of Döring (1936), which give for the mean value of λ (at 290° K.) -3.0×10^{-5} , and for $d\lambda/dT + 2.5 \times 10^{-8}$. Putting $q=9/2$, and expressing Z in kg. mm.⁻², this gives

$$k_{\text{eff}}/k = 1 + 0.27Z, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5.16)$$

$$b_{\text{eff}}/b = (1 + 0.015Z)/(1 + 0.27Z). \quad . \quad . \quad . \quad . \quad . \quad (5.17)$$

Even for fairly small internal stresses (of the order, say, of 1 kg. mm.⁻²), the increase in k_{eff} , and the change in the form of the magnetization curve may be relatively large, but, as shown by (5.17), the effective rotational magneto-caloric coefficient decreases.

This treatment, though crude, shows in a semi-quantitative way how internal strains may result in magnetization curves and magneto-caloric temperature changes very different from those estimated for undeformed polycrystalline material. An elaboration of the treatment is obviously possible, but it hardly falls within the scope of the present paper; such elaboration, moreover, would be unlikely to be quantitatively effective unless carried out with direct reference to particular materials for which adequate experimental data were available.

Atomic theory.

The problem confronting atomic theory, as distinct from any statistical theory of stress distribution, and of grain orientation in polycrystalline material, is the explanation of the order of magnitude of the crystal anisotropy and magnetostriction coefficients. A satisfactory quantitative solution is far from having been reached. The complexity of the problem is apparent from a detailed paper by Van Vleck (1937), who gives full references to earlier work. Van Vleck's treatment is on a Heitler-London, Heisenberg basis. A reformulation on a collective, or itinerant electron basis is given by Brooks (1940), and there is also a recent paper on cobalt by Vonsovsky (1940). The effect of purely magnetic coupling between the spins is negligible, and exchange interaction is essentially isotropic in respect of the direction of magnetization. To account for anisotropy, interactions formally similar to those between dipoles and quadrupoles are required. Van Vleck shows that such interaction is simulated by spin-orbit coupling combined with orbital valence coupling, and although no detailed assumptions are made about the orbital states ("what the orbital states look like") he further shows, by a somewhat elaborate mathematical treatment, that this can give effects of the required order of magnitude.

In ferromagnetic metals the gyromagnetic ratio shows that the orbital moments are almost completely "quenched", that is, the energy distribution of orbital states is such that a magnetic field produces little change in the total resultant orbital magnetic moment; in somewhat crude terms, the orbital moments are strongly constrained to particular directions relative to the crystal axes. Even with weak spin-orbit coupling, however, changes in spin orientation may be accompanied by orbital changes with associated change in the orbit-orbit interactions, which depend not only on the relative orientation of orbital moments but also on their orientation relative to the line joining the atoms.

The magneto-caloric effects associated with rotational processes may be considered in much the same way as those associated with changes in intrinsic magnetization. As the magnetization vector rotates from an easy to a difficult direction, the effective reverse field increases, and the intrinsic magnetization decreases, with an increase of entropy. Numerical consideration of the data shows that this spin entropy increase alone would amount to less than one-tenth of the total increase. The major

Although it is well known that the initial susceptibility in general increases with temperature to a sharp maximum at or near the Curie point, comparatively few recent systematic investigations of the precise temperature variation seem to have been made. Among them the most extensive is that of Kahan (1938). Some values derived from his results are given in Table VI., but it should be stated that owing to the rather wide spacing of the numerical data for κ_0 as a function of temperature, as well as the irregularities, there is a large probable error in the derived values of $d\kappa_0/dT$, and consequently also in the coefficients c and b' .

The initial susceptibility is strongly structure-sensitive, being markedly dependent on thermal and mechanical treatment, even more than on degree of purity. The two specimens of iron (Armco and Yensen) were of high purity, but had different thermal treatments. The nickel specimen

TABLE VI.

Initial susceptibility and related quantities,
derived from data of Kahan (1938).

$T=290^\circ \text{K.}$

κ_0 , initial susceptibility.

$c = -\frac{1}{2}T(d\kappa_0/dT).$

$b' = -(T/\kappa_0)(d\kappa_0/dT).$

	κ_0	$d\kappa_0/dT$	$-c$	$-b'$
Fe (1)	17.1	0.07 ₅	10.9	1.2 ₃
Fe (2)	10.3	0.03 ₄	4.9	0.9 ₆
Ni	12.2	0.05	7.2	1.2
Co	2.3	0.009	1.3	1.1

(a Mond nickel, 99.25 Ni) after other heat treatments, had initial susceptibilities of 10.9 and 2.06, with temperature coefficients of about 0.065, and 0.007. Another Mond nickel (99.71 Ni) studied by Kirkham (1937) gave $\kappa_0=26.1$, $d\kappa_0/dT=0.20$. The original papers must be consulted for fuller details of the materials. It is evident that the values are characteristic of the particular specimens investigated, and that Table VI. is not to be regarded as characterizing iron, nickel and cobalt in the same sense as Tables I., II. and III. Data useful in interpreting magneto-caloric effects must, at least in the present connection, be obtained for the same specimens as those used in the magneto-caloric investigations. None the less, Table VI. is useful in giving some idea of orders of magnitude. In the approximately reversible range, which seldom extends over more than a few oersteds, from (6.2), using the c values in the Table, the cooling effect will usually be too small for accurate measurement, though it might contribute appreciably to the total thermal change over a wider field range. Comparison of the values of a' (equal to $\frac{1}{2}b'$, where b' is given in Table VI.) with those of a in Table V. suggests that the initial cooling associated with

reversible boundary movement would normally considerably exceed the heating associated with increase of intrinsic magnetization. (It is the resultant of the two effects which corresponds to the coefficient a' .) Comparison of b' of Table VI. with b of Table V. shows that the thermal effect is of the same sign as that associated with rotation, and of comparable magnitude. As far as the magneto-caloric effect is concerned, therefore, reversible boundary movement is formally covered by the second term in the general equation (5.11), but with a modified value of the coefficient b .

Mechanism and atomic theory.

In so far as reversible boundary movement takes place without change in boundary area and effective thickness, the change in internal energy and in entropy corresponds to the difference of these quantities in the final and the initial state in the region traversed by the boundary. For a 180° boundary the direction of magnetization is simply reversed and there would normally be no associated thermal effect other than that due to the intrinsic magnetization change, already covered by the first term of (5.11). This would be true also of 90° boundaries unless there were some differentiation between the easy axes, which are strictly equivalent only in an ideal limit. This differentiation may be produced by local strains. These strains are set up as a ferromagnetic cools down from a high temperature, and the intrinsic magnetization, developed as the material cools from the Curie point, takes up a direction energetically favoured by the local strain and crystalline anisotropy. With the small strains in well annealed materials these directions will lie along or nearly along one of the crystal easy axes. In the absence of strains of purely mechanical origin, there are still small strains set up by the magnetostrictive changes with developing magnetization as the material cools, a process shown by Kersten (1931) to account almost quantitatively for the maximum observed initial susceptibilities of many materials. It may be said that in general the domain axis along which the local magnetization vector lies in the "natural" zero field state of a ferromagnetic material (ordinarily, but not necessarily, a demagnetized state) is energetically favoured to a greater or less extent over other ideally equivalent easy axes. Although 180° boundary movements may contribute largely to the initial susceptibility (the movement being controlled by surface or internal free pole effects), the thermal changes on adiabatic magnetization must be mainly due to the movement of 90° boundaries. The atomic theory of the effect is essentially the same as that discussed in connection with rotational processes (§ 5). It should, perhaps, be explicitly stated that 90° boundary movements can hardly be involved in the magnetization of cobalt. Data on this remarkable material are unfortunately very sparse, but the measured values of the initial susceptibility are invariably small compared with those for iron or nickel, and the thermal effects may be mainly due to rotations.

Boundary formation.

Mention must finally be made of the formation and dissolution of domain boundaries, processes which are not necessarily irreversible. The whole question is complex and speculative, and no more than tentative suggestions can be made here. As the field is reduced from a high value at which approximate saturation is attained, boundaries must be gradually formed, at first of low energy, as they separate regions in which the directions of the intrinsic magnetization are almost parallel. These domain boundaries may at first run along grain boundaries, but as the bulk magnetization decreases, and the directions of the intrinsic magnetization diverge more and more, the domain boundary pattern (built up in such a way as to decrease the energy which would otherwise arise from external and internal demagnetizing effects) becomes more and more complex. In so far as a particular pattern is uniquely associated with a particular field (and state of bulk magnetization) the process is reversible. Boundary formation necessarily involves an increase in internal energy and in entropy, and the thermal effects will be of the same sign as those associated with decrease in intrinsic magnetization. Boundary formation and dissolution will therefore increase numerically the effective value of the intrinsic magnetization magneto-caloric coefficient α (Table V.). It would hardly be profitable to attempt a "theoretical" calculation of the magnitude of the effect, but it should be kept in mind in connection with any analysis of the magneto-caloric data.

§ 7. IRREVERSIBLE PROCESSES.

Elementary irreversible processes.

Under certain conditions irreversible translational movement of a domain boundary may take place. After passing through a succession of equilibrium positions as the field is gradually increased the boundary may, at a critical value of the field, move spontaneously, through a succession of non-equilibrium positions, to a new position of equilibrium, after which the reversible movement may be resumed as the field is further increased. Such a process is, in general, irreversible, and the critical field values may be widely different for increasing and decreasing field. The general mechanism of domain boundary movements has been investigated by Becker and others, the irreversible movements, in relation to hysteresis and coercivity, having been treated particularly by Kersten. In the earlier work consideration was given mainly to the effect of localized inhomogeneities of strain in determining the variation with position of the energy of a boundary. The treatment is presented in detail by Becker and Döring (1939, pp. 176–217), and several shorter accounts have appeared (*e.g.* Stoner 1944, 1948). More recently Kersten (1943a, b) has examined the effect of non-magnetic inclusions on boundary energy, and he again obtains good agreement with experiment for a wide range of materials, but without the necessity for postulating intrinsically improbable

types of stress distribution. Both the "strain" and "inclusion" treatments have, however, been strongly criticized by Néel (1946), who concludes that the agreement with experiment which is claimed has been achieved only through the implicit assumption of an extremely regular distribution of strain or inclusion centres; with a random distribution, the mechanism could not account for coercivities of more than a fraction of an oersted. In Néel's view, the inhomogeneities in strain or composition give rise to fluctuations in the direction and intensity of the intrinsic magnetization throughout the volume of domains, and these, through the consequent appearance of internal free poles (non-zero values of the divergence of the intrinsic magnetization vector), give rise to a magnetic energy with a volume rather than a surface distribution. For two neighbouring domains the magnitude of this magnetic energy varies with the position of the boundary between them. It is shown, by a rough method, that the observed orders of magnitude of coercivities can be accounted for in this way, and although it would be extremely difficult to develop the theory in a rigorous quantitative manner, there can be little doubt as to the validity of Néel's general treatment, or of his criticism of the earlier treatments. It is to be noticed that although there is a fundamental difference in the two types of treatment with regard to the "localization" of the energy—with Kersten it is in the boundary zone itself, with Néel in the surrounding volume—in both the variation of energy with boundary position is determinative of whether changes in magnetization with changing field occur reversibly or irreversibly.

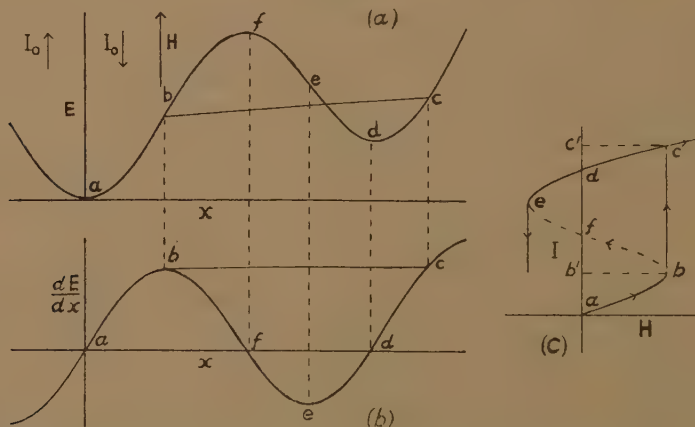
The possibility of rotational hysteresis was envisaged by Akulov (1933) but it has usually been regarded as having little direct bearing on actual hysteresis effects, as the critical fields for rotational "jumps" are much higher than the fields at which the same directional changes are normally brought about by boundary movements. In particles below the critical size (depending on shape) for which boundary formation is energetically possible, change of direction of intrinsic magnetization must, however, take place entirely through rotational processes, and rotational hysteresis effects may be fully manifested. It was suggested independently by Néel (1947) and by Stoner and Wohlfarth (1947) that the very high coercivities observed in certain classes of ferromagnetic material could be explained on this basis. A detailed quantitative treatment of the variation of magnetization with field for single-domain, anisotropic particles, and of the behaviour of aggregates of such particles, has been given by Stoner and Wohlfarth (1948), and its bearing on the properties of permanent magnet alloys of the dispersion-hardening type and other materials is fully discussed. The formal mathematical treatment applies equally to magneto-crystalline anisotropy, strain anisotropy, and shape anisotropy. This particular mechanism of hysteresis is unlikely to play an important part in most of the materials whose magneto-caloric behaviour has been examined, and it need not be considered here in detail. There is, however, one relevant general point which may be noted. In

this case the detailed mechanism of the magnetization process is known, and numerical details of the behaviour have been worked out. By an extension of the treatment, which has been carried out by Rhodes (1948), precise theoretical calculations may be made of the distribution along the magnetization curve of the irreversible heating accompanying adiabatic magnetization. Though this may be a unique example of a rigorous calculation of an irreversible magneto-thermal effect, it affords a practical demonstration that irreversible as well as reversible effects are amenable to precise theoretical treatment if sufficient is known about the mechanism of the processes involved.

General character of irreversible processes.

The general character of the energy relations in irreversible processes may most simply be described by reference to an idealized example of domain boundary movement, for which the relevant curves are shown in fig. 5.

Fig. 5.



Diagrams illustrating reversible and irreversible boundary movements.

The boundary is parallel to the YZ plane, with its central plane passing through the energy minimum at a in zero field. As the field increases the boundary moves in the direction of increasing x . Corresponding points in (a), (b) and (c) are marked by the same letters.

It is supposed that the internal energy varies with the position of a 180° boundary in the way indicated in fig. 5(a), the intrinsic magnetization being directed upwards (the $+y$ direction) to the left, and downwards to the right of the boundary, and that the field is applied in the upward direction. For zero field the boundary is at the energy minimum at the origin. In a field H the total energy is given by $E - 2HI_0x$ (using the symbols in a "diagrammatic" rather than a strict sense), and as the field increases the boundary moves to the right with the minimum of the total energy curve, taking up a position for which $(dE/dx) - 2HI_0$ is zero. This corresponds to a point on the (dE/dx) curve (fig. 5(b)) at which it is

intersected by a line parallel to the x axis at a height proportional to H . It will be clear from the curves that the movement is reversible up to b , but that at b the boundary moves "spontaneously" (*i. e.* without change of the field) to a new position of equilibrium c , the process corresponding to a positive jump in the magnetization (fig. 5(c)). On decreasing the field the path cde is followed, a negative jump occurring at e .

In an adiabatic change, the increase in internal energy in the positive jump, namely $\int H dI$, is given by the rectangular area $b'bcc'$ (fig. 5(c)). If the process could be carried out isentropically, by appropriate adjustment of the field at each stage, the path $bfedc$ would be followed. The portion bfe is essentially unstable, corresponding to a succession of total energy maxima, but, in an idealized sense, this process is reversible. The heat developed irreversibly in the jump process (proportional to the entropy increase in an irreversible adiabatic change) corresponds to the difference between the integrals $\int H dI$ along the paths bc and $bfedc$. This last integral (made up of the two positive parts $b'bf$ and $c'cd$ and the negative part fed) is equal simply to the difference in internal energy in the states b and c , as given by the difference in height of b and c in fig. 5(a).

From the knowledge involved in the curves of figs. 5(a) and (b) of the mechanism of the elementary process, the corresponding elementary magnetization curve, fig. 5(c), can be uniquely derived; the irreversible heating effects can be calculated, and the difference in the non-thermal part of the internal energy in states such as b and c is known. The converse procedure is not possible; given, for example, the part $abcde$, of an elementary magnetization curve, it is not possible to determine, from that information alone, the difference in the non-thermal or "magnetic" internal energy in states corresponding to the points b and c , or a and d . In this particular case this energy is greater at c than at b , and greater at d than at a ; but no generally valid *a priori* statement can be made as to the difference in the internal magnetic energy for points on an initial magnetization curve and points corresponding to the same field on the upward and downward branches of the associated hysteresis loop. The internal energy of a ferromagnetic at remanence, for example, may be equal to, greater than, or less than that in the demagnetized state. Differences in internal "magnetic" energy between any two points on a magnetization curve can, however, be unambiguously determined when the magnetic measurements are supplemented by magneto-caloric measurements (see (2.14)), whether or not the changes are reversible.

Curves or tables for the variation of internal magnetic energy along a magnetization curve are re-expressions of the magnetic and magneto-caloric results rather than interpretations of them. Little progress in interpretation in terms of the elementary processes occurring can be made unless information is also obtained about the variation of magnetic entropy, which is possible when the changes are reversible. Over a wide range of an ordinary magnetization curve reversible and irreversible changes are superposed, and a pre-requisite of any satisfactory analysis of

the results is an unambiguous separation of the associated reversible and irreversible heat changes. A satisfactory separation has not hitherto been made, nor are data available from which it could be made, and for this reason no more than a partial interpretation of any magneto-caloric results can at present be given (see § 8). As the separation is essential not only for the interpretation of the reversible changes over the whole of a curve, but also for the sharper characterization of the irreversible changes, the question of how it might be effected must be briefly discussed.

Separation of reversible and irreversible effects.

The heat developed in a hysteresis cycle can be determined with good accuracy from the area of the hysteresis loop, which should be equal to the aggregate heat developed as determined "calorimetrically". If the loop is symmetrical, moreover, the heat developed in the change of field from $-H_m$ to $+H_m$ is the same as for the change from $+H_m$ to $-H_m$, and equal to half the total heat developed in the complete cycle. This is the total of the information about the irreversible heat developed which can be obtained, without special assumptions, from a consideration of the magnetization curves (except, of course, for the obvious fact that in regions where the change of magnetization with field is strictly reversible there is no irreversible development of heat). There is, for example, no general relation between the irreversible heat development along an initial curve (0 to H_m) and in the complete cycle ($\pm H_m$). The differentiation of the reversible and irreversible heat changes throughout the course of a magnetization curve is very difficult, and seems to have been attempted only by Okamura (1936), who boldly gives curves and tables for each of the two types of change. The principle of the method seems not entirely unsound for giving the separation "in a good approximation", which is all that is claimed.

The essential assumption underlying Okamura's method is that the heat developed irreversibly, $\Delta Q'_{ir}$, when the field is changed from, say, $+H_2$ to $+H_1$ is the same as that when the change is from $+H_1$ to $+H_2$, and that the heats developed reversibly, $\Delta Q'_r$, are equal and opposite. If the changes in H were small, this would be a reasonable assumption, but an accurate measurement of the heat change is usually impracticable unless the field change is relatively large. In practice very large changes in H were made (*e. g.* for nickel between $+235$ and values of H ranging from $+190$ to -240). There is no guarantee that the subsidiary hysteresis loops were closed (there is no information on this point), and if a cyclic condition were established for each of these subsidiary loops, neither the descending nor the ascending branch of these loops would, in general, coincide with the main curve primarily under investigation. For these reasons, quite apart from any question of the general accuracy of the results, Okamura's curves cannot be accepted as giving more than a very rough qualitative indication of the trend of the separate heat changes. In some cases the indication may, by accident, be reasonably accurate.

quantitatively, but it is virtually impossible to pick out such favourable cases on the evidence of the paper alone.

Since it is essential for a proper understanding of the magneto-caloric effects that an unambiguous separation of the reversible and irreversible heat changes should be made, it is perhaps appropriate to indicate a method by which this might be done. It is first necessary to determine the differential and reversible susceptibility along the course of the magnetization curve under investigation. The total change of I over any range can then be represented as the sum of two parts, one corresponding to the change resulting from reversible, the other from irreversible, processes, and separate I_r, H and I_{ir}, H curves may be obtained. From I_r, H curves at different temperatures (including at least one above and one below the temperature at which the magneto-caloric measurements are made) the adiabatic temperature change associated with the reversible processes can then be unambiguously determined from the basic thermodynamic equation (2.8), namely

$$\left(\frac{\partial T}{\partial H}\right)_s = -\frac{T}{C_p} \left(\frac{\partial I}{\partial T}\right)_H \quad \dots \quad (7.1)$$

The mutual consistency of magneto-caloric results and those on the reversible susceptibilities can be checked by a comparison of the directly determined and the indirectly calculated thermal changes in strictly reversible regions of the magnetization curve. For other parts of the curve the irreversible heating is found from the difference in the total heating and the calculated reversible heating. There would, of course, be many points of detail to be considered in the experimental work if reliable results were to be obtained, but measurements of reversible susceptibility, using small field changes, are entirely practicable. (Work in progress (1948) along these lines, but in a different connection, in the Leeds University Physics Department by Mr. W. D. Corner show that with care, but without undue elaboration of apparatus, consistent results of considerable precision may be obtained. The values of the reversible magneto-caloric heat changes obtained indirectly would have a reliability at least as high as those obtained from the direct measurements.)

The magneto-caloric and the reversible susceptibility measurements should be made on the same specimens, or, if there are technical difficulties in attaining that ideal, owing, for example, to a requirement for specimens of different shapes and sizes, on specimens as nearly as possible of magnetically identical material. The procedure proposed has many advantages over that adopted by Okamura, but there may well be other more convenient procedures which are equally effective for the purpose in view. It seems quite clear, however, that suitably designed supplementary experiments must be carried out in conjunction with those on the magneto-caloric effect if a full interpretation is to be obtained of the results, and, even more important, if the fullest amount of significant new information is to be derived from them about magnetization processes in ferromagnetics.

§ 8. DISCUSSION OF EXPERIMENTAL RESULTS.

General.

Measurements of the adiabatic temperature changes accompanying magnetization in low and moderate fields have been made by Adelsberger (1927), Constant (1928), Honda, Ôkubo and Hirone (1929), Ellwood (1930), Townsend (1935), Okamura (1936), Hardy and Quimby (1938), and Bates with Weston (1941), Healey (1943), Edmondson (1947) and Harrison (1948a, b). There is much of interest in the earlier work, but almost all the results are included in, or have been superseded by, those of Okamura, Hardy and Quimby, and Bates, from whose papers the illustrative examples will be taken. The original papers must be consulted for details of the experimental arrangements, though it should be mentioned that Okamura used a multiple specimen test sample, while single specimens in the form of rods or stout wires were used by Hardy and Quimby and by Bates. The overall sensitivity obtained by the different workers is much the same, ranging from about 200 to 900 erg. cm.⁻³ per scale division.

Okamura's work is very extensive, and he gives full details in both graphical and tabular form of results on nine different materials, including Armco iron, nickel, cobalt and various steels; but practically no details are given of the composition of the specimens. An objective test of the reliability of the results is provided by a comparison of the hysteresis loss as determined from the heat developed in traversing a cycle, and from the area of the hysteresis loop. Okamura's own table (p. 790) shows in some cases relative differences of up to about 80 per cent, and absolute differences of as much as 7000 erg. cm.⁻³. For this and other reasons (which soon become apparent when a critical examination of particular sets of results is made) it must be concluded that Okamura's results, though they have value as a preliminary experimental survey of a wide field, cannot in general be accepted as quantitatively reliable. This is the more unfortunate as an enormous amount of work has been carried out (including supplementary experiments on the thermodynamic relations), and the results are given so fully.

In the work of Hardy and Quimby there is satisfactory agreement between the two estimates of the hysteresis loss, the differences, according to their table (p. 320), usually being within 3 per cent, except for a low loss material for which the 30 per cent difference corresponds to an absolute difference of only 200 erg cm.⁻³. Bates (1941, p. 29), in his study of nickel, obtains even closer agreement. Hardy and Quimby made measurements on annealed and unannealed nickel and Armco iron, and on a carbon steel, and give the full results in tabular as well as graphical form. Bates and his collaborators have examined a variety of specimens of nickel (1941), iron (1943, 1948b), and cobalt (1947) and some seven ferromagnetic alloys (1948a). In a few cases full numerical results are given. In most cases, however, the results are given only in the form of graphs (usually of Q' and of $\int H dI$ against I). These show admirably the general character of the changes, but do not lend themselves to any but the roughest numerical analysis.

Method of analysis of results.

The method of analysis which it is appropriate to adopt depends on the information available additional to that provided by the magneto-caloric measurements themselves. If no other information were available, no "analysis" of the magneto-caloric results would be possible, and in some cases this is effectively the position. The magnetization curve itself (which, incidentally, is not always given with published results on the magneto-caloric effect) does, of course, give other information, but in this paper the central problem under consideration is not so much the direct analysis of the magnetization curve as the question of how far the magneto-caloric results can contribute to that analysis, that is, to the determination of the extent to which the various possible processes are involved in the change of magnetization over any section of the curve. The most serious obstacle in the way of a complete analysis at present is the absence of reliable information as to the reversible and irreversible heat changes separately (see § 7). In certain cases, the contribution to the heat changes from changes in intrinsic magnetization can be estimated with fair precision (§ 4), and in other cases the experimental results required for the estimate could be obtained without undue difficulty. It seems appropriate, therefore, to subtract this part of the heat change from the observed total, and to consider the remainder. The basic equation to be used in considering the experimental results may be written

$$Q'' = Q' - a \int d(IH) = \int b'' H dI, \quad (8.1)$$

where Q' is the observed heat developed measured from an appropriate initial state to which correspond the lower limits of the two integrals. This equation is intended to recall (5.11); the dropping of Δ , in the interests of economy of symbols, should cause no confusion. The coefficient a , the magneto-caloric coefficient for change of intrinsic magnetization, has the same significance, but b'' is to be regarded as a possibly variable coefficient whose value is to be determined from the experimental results. Formally, the equation for b'' may be written

$$b'' = (1/H)(dQ''/dI). \quad (8.2)$$

Strictly the derivative should be written in partial form, but it may be understood that it is to be evaluated from the experimental results which correspond to adiabatic conditions. Apart from apparent, but irrelevant, objections to the use of (8.2) in regions where discontinuous changes occur, particularly at or near zero values of H , the equation cannot in practice be used in this form owing to the usually small number, the often very irregular spacing, and the invariable "scatter" of the experimental values, and it must be replaced by the approximate equation

$$b'' = \Delta Q'' / \Delta \int H dI. \quad (8.3)$$

No useful purpose would be served by attempting to formulate general

rules for the choice of intervals in evaluating these mean values of b'' , for the choice is largely dictated by the extent and character (in particular the regularity) of each particular set of experimental results.

The order of magnitude of the experimental coefficient b'' over any range of a magnetization curve, and its variation over the curve as a whole, are to be considered in the light of the theoretical treatment in the preceding sections. For a homogeneous polycrystalline specimen, free from internal strains, it would be expected that over the range from technical saturation to remanence b'' would approximate to the value of b , corresponding to rotation of the magnetization vector in the magneto-crystalline field, and of which estimates are given in Table V. Even assuming complete reliability of the magneto-caloric results, no close agreement is to be expected, even for well annealed material, for the estimates themselves are subject to considerable uncertainties, and the modification in b due to quite small internal strains may be large. More generally, the magnitude and variation of b'' over this part of the curve may be expected to correspond to those of the theoretical b_{eff} of (5.15), which formally takes account of strain effects. (In view of the difficulties in making an accurate *a priori* estimate of b_{eff} , it is useless at present to attempt to draw any conclusions from the experimental results about thermal effects associated with the formation or dissolution of domain boundary patterns.) Over irreversible parts of a magnetization curve, in which discontinuous magnetization processes occur, relatively large changes in b'' are to be expected, reflecting the extent of the irreversible development of heat.

The general validity of the theoretical treatment (or of the reliability of the experimental results, according to the point of view) is to be assessed first from the similarity in the quasi-symmetrical relation between the Q'' , H and the $\int H dI$, H curves (making allowances for the irreversible effects) for materials whose Q' , H curves are entirely different, and secondly from the reasonableness or otherwise of the values of b'' considered in the light of the theoretical discussion.

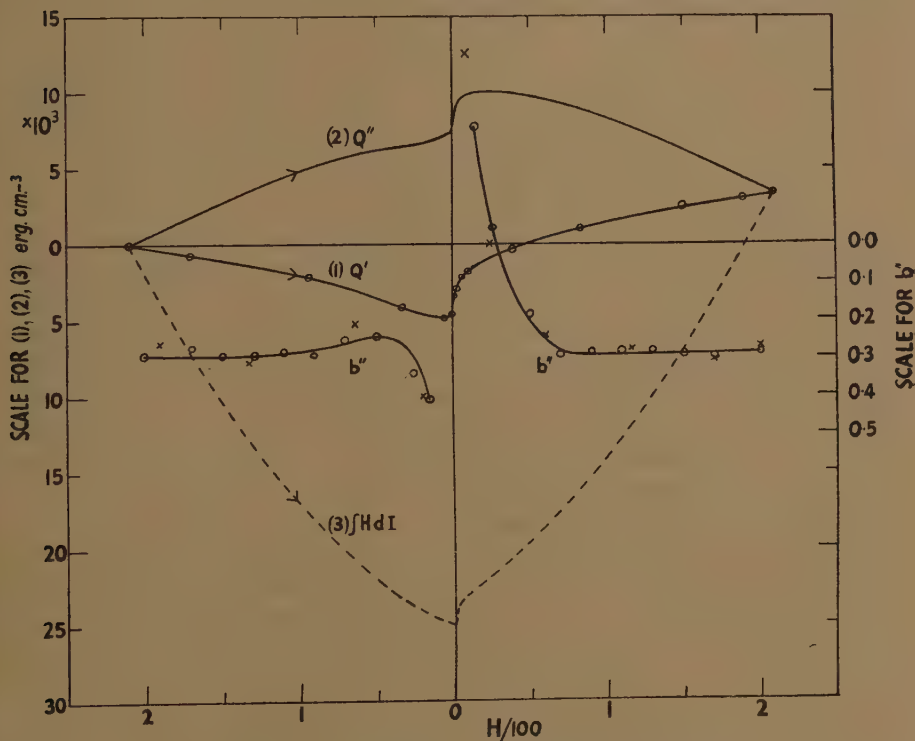
Many sets of experimental results, in the form of numerical data or curves, have been examined by the method outlined above. It is, however, only in few cases that the magneto-caloric data are sufficiently detailed for more than a rather superficial analysis. Little purpose would be served in discussing in detail all the more reliable results (such as those of Bates, and of Hardy and Quimby) for it is hoped that experimental results additional to those on the magneto-caloric effect itself will be obtained later which will make possible more complete and satisfying analyses. The few examples given below will be sufficient to indicate the general method and the character of the conclusions which can at present be obtained.

Iron.

The results for the heat developed under adiabatic conditions in annealed Armco iron, as obtained by Hardy and Quimby (1938, Table II. c,

fig. 4) are shown graphically in the Q' curve of fig. 6. The curve Q'' , representing the heat developed other than that due to changes in intrinsic magnetization, is obtained from (8.1). The temperature is given as "about 41.3°C ," say 314°K ., for which the estimated value of a is 3.8×10^{-2} , slightly greater than the value given in Table V. for 290°K . The values derived for b'' (see (8.3)) using directly the tabulated values of Q' (shown by circles on the Q' curve) are indicated by crosses. In deriving values of b'' in this way, however, the effect of unsystematic experimental errors is

Fig. 6.



Magneto-caloric effect in annealed Armco iron.

Data for curves (1) and (3) from Hardy and Quimby (1938).

Q' , heat developed.

Q'' , heat developed less that due to change in intrinsic magnetization.

$b'' = \Delta Q'' / \Delta \int H dI$. Circles, values derived from smoothed curves.

Crosses, values from experimental points.

greatly magnified, and a more significant set of values is obtained by taking values for $\Delta Q''$ and $\Delta \int H dI$, for suitable intervals, from large scale curves, corresponding to (2) and (3), drawn smoothly between the experimental points. Values obtained in this way are shown by circles, and the b'' curve is drawn smoothly between them.

As the field is increased from $-H_m$, b'' is approximately constant until the remanence point (corresponding to $H=0$) is approached, when it becomes increasingly negative. In passing through the zero field region, there is a rapid change in the value of b'' , which, after passing a maximum (in this case positive), gradually decreases to the same approximately constant value as over most of the negative field range. The general character of the variation of b'' in the low field region is clearly related to the irreversible development of heat, which is most marked along the steepest part of the magnetization curve. The way in which the presence of irreversible heat development is manifested in the b'' curve is very striking, but for reasons already discussed (§ 7) no useful purpose would be served at present by considering the low field region in detail. The mean value of b'' for the range from $-H_m$ to 0 (*i. e.* from $-I_{\max}$ to $-I_{\text{rem}}$), over which the magnetization change is mainly reversible, is -0.30 . The estimated value of b for polycrystalline iron completely free from strain at the temperature of the experiments is -0.63 (slightly higher numerically than the value in Table V.). Bearing in mind that the estimated value is very uncertain (a different weighting of the sparse experimental data available could well lead to an estimate some 40 per cent lower), and that, owing to the method of analysis, the uncertainties in the magneto-caloric data are accumulated in the derived values of b'' , the order of agreement in absolute magnitude must, at this stage, be regarded as more remarkable than the numerical discrepancy. At the same time an illustration is given of the impossibility of extracting fully the information inherent in the magneto-caloric results unless reliable and precise data are available for the basic characteristics of the material.

The results of Bates on annealed iron agree in general character with those of Hardy and Quimby, but they are given only in graphical form; a rough estimate of the mean value of b'' for the range from I_{\max} to I_{rem} may, however, be made. For Armco iron (99.89 Fe) (Bates and Harrison 1948 b, fig. 5), $b'' = -0.26$; for Hilger electrolytic iron (99.96 Fe), (*loc. cit.*, fig. 9), $b'' = -0.20$. Some older results of Bates and Healey (1943, fig. 7) give for the same Armco iron $b'' = -0.22$. (These values are derived from the results for the largest cycles, with H_{\max} between 350 and 400. Space precludes an adequate discussion of the interesting results for smaller cycles; the indications are that the numerical values of b'' are somewhat greater). These values agree well with each other, and as to order of magnitude with the estimated value of -0.58 for b (Table V.).

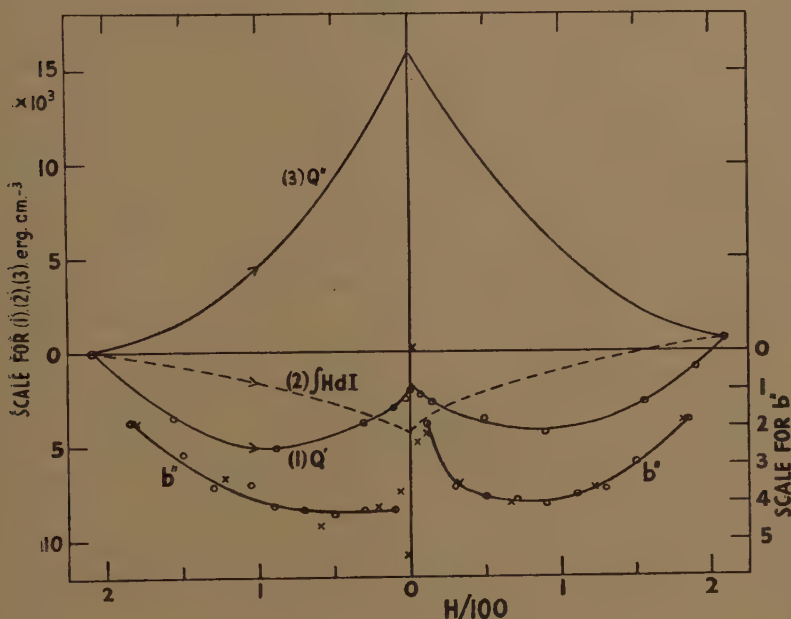
For unannealed Armco iron, the mean values of b'' are -0.38 (Hardy and Quimby 1938, Table II. *d*) and -0.36 (Bates and Healey 1943, fig. 1), and the results of Hardy and Quimby (Table II. *e*) give practically the same value for a carbon steel (1.08 C.). These values are not very different from those for the annealed materials, as is in accordance with the general conclusions about the effect of internal strains in iron (§ 5), but they appear to be systematically greater; in the absence of any reliable data for the temperature coefficient of the magnetostriction of iron, the possible significance of this cannot be profitably discussed.

It should perhaps be noted that for Armco iron, Okamura (1936, Table VII. c, fig. 13 b), for a field range practically the same as that in fig. 6, records a heat change some six times as great as Hardy and Quimby, and of opposite sign (*i. e.* a heating rather than a cooling). This, as suggested by Hardy and Quimby, must be due to "unfortunate flaws in his experimental method". A curve drawn from Okamura's data for Armco iron is given by Becker and Döring (1939, p. 268).

Nickel.

Results for annealed nickel (about 99.4 Ni) obtained by Hardy and Quimby (1938, Table II. a, fig. 4) are shown in fig. 7. The Q'' and b'' curves

Fig. 7.



Magneto-caloric effect in annealed nickel.

Data for curves (1) and (3) from Hardy and Quimby (1938).

For symbols, see fig. 6.

are obtained in the same way as for iron. (For the coefficient a the value 18.4×10^{-2} was adopted for the temperature 314°K. , as compared with 14.3×10^{-2} in Table II. for 290°K.)

Over the low field region in which most of the irreversible changes occur the variation of b'' is of the same character as for iron, and is similarly explained. There is, however, in contrast to iron, a definite decrease in the numerical value of b'' in the high field regions. None the less, the mean value of b'' over the range from $H = -210$ to $H = 0$, equal to -3.7 , shows the same order of agreement as for iron with the estimated value of -5.9 at this temperature for polycrystalline material free from internal strains. The primary fact to be noticed is that the large difference in

the estimated values of b for iron and nickel (-0.6 and -5.9 at 314°K.) is paralleled by the difference in the mean values of b'' derived from the magneto-caloric effect for annealed materials (-0.3 and -3.7 for the materials of figs. 6 and 7) over regions in which the greater part of the change is reversible. The numerical decrease in b'' in higher fields for nickel may tentatively be attributed to the effect of residual strains which have not been removed in the annealing process. As shown in § 5, in nickel, which is most sensitive to stresses, the effect is to decrease the effective value of b (as is confirmed by the results on unannealed nickel referred to below) and to increase the effective anisotropy coefficient. The effective anisotropy will vary from grain to grain in the polycrystalline material, and as the field increases the contribution to the change of magnetization from the more strongly anisotropic grains will become relatively more important, with a consequent falling off in the value of b'' .

The most extensive investigation of nickel is that of Bates and Weston (1941), who made measurements on a number of high purity materials (99.53 to 99.98 Ni) in both the hard drawn and annealed states, and also examined the effect of applied loads. The results are mostly given only in graphical form (usually Q' , I and $\int H dI$, I curves). The curves given agree in general form with those of Hardy and Quimby, but reliable estimates can only be made of the mean value of b'' from I_{\max} to I_{rem} . For four annealed specimens, with a maximum field of about 200 (*loc. cit.*, figs. 8, 10, 12, 15), the estimated values of b'' are -4.2 , -4.3 , -3.7 and -4.2 . These values agree very closely with the estimated value of -4.5 for b (Table V.), and suggest that, after allowing for the effect of change of intrinsic magnetization, most of the remaining effect is due to rotation of the magnetization vector in the natural anisotropic crystalline field. For one of the specimens, however, a larger cycle was examined, with a maximum field of 411 (*loc. cit.*, fig. 16). For this the mean value of b'' is numerically much lower, about -1.4 ; this is due to a large cooling effect as the field is changed from -411 to about -200 (which corresponds to a change of I from about -455 to -430). This initial cooling appears to be comparable with, and possibly larger than, the maximum cooling which can be attributed to the change in intrinsic magnetization, and if confirmed may be an indication of boundary formation effects. It may be noted that this specimen was given a somewhat unusual annealing treatment (*loc. cit.*, p. 22).

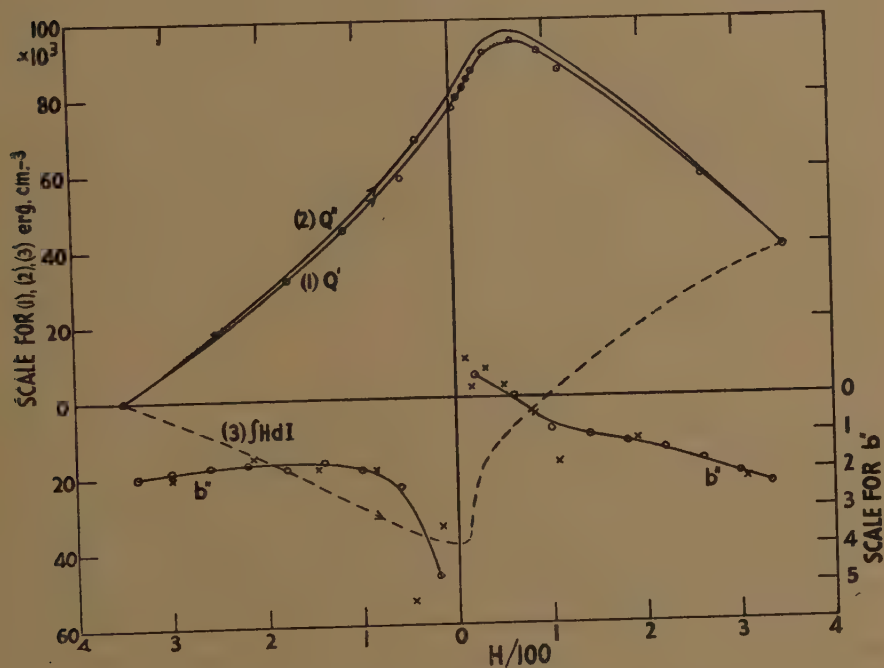
For hard drawn nickel the results of Bates (*loc. cit.*, figs. 5, 7) give mean b'' values of about -0.65 and -0.84 . These lower numerical values (which are also found from the results for nickel under tension) are qualitatively in accordance with the theoretical treatment of § 5. The results of Hardy and Quimby, however, for the material of fig. 7, but unannealed (*loc. cit.*, Table I. b), give, not only a positive mean value for b'' of 0.37 for the range from I_{\max} to I_{rem} , but also values which are positive for each (experimental) interval over the whole curve. There is no obvious means of accounting for a positive value of b'' over the whole of

the curve in terms of the magnetization processes which have been considered, and since the recorded cooling is some four times as great as that found by Bates for similar cycles and presumably similar materials, it is probably inadvisable to speculate on the theoretical interpretation of this particular set of results of Hardy and Quimby until confirmatory experiments have been carried out.

Cobalt.

A very thorough study of cobalt (98.4 Co) has been made by Bates and Edmondson (1947, 48), and full numerical details of the results are given. Their results for an annealed specimen (*loc. cit.*, fig. 10, Table VII.) are shown in curves (1) and (3) of fig. 8.

Fig. 8.



Magneto-caloric effect in annealed cobalt.

Data for curves (1) and (3) from Bates and Edmondson (1947).

For symbols, see fig. 6.

Owing to the relative smallness of the effect of change of intrinsic magnetization in cobalt the Q'' curve differs very little from the Q' curve. The striking difference of the Q' curve from those for iron and nickel is due to the fact that for cobalt almost the whole of the reversible adiabatic temperature change is due to rotation of the magnetization vector (*cf.* fig. 4). The variation of b'' in the low field region is similar to that for iron and nickel, but the effect of irreversible changes persists over a wider

field range. (It may be noted that the aggregate irreversible heating, measured by the value of Q' at $+H_m$, is much greater for the cobalt of fig. 8 than for the iron and nickel of figs. 6 and 7.) Over most of the range from $-I_{\max}$ to $-I_{\text{rem}}$, b'' varies only slightly. The mean value for this range is -2.0 , to be compared with the uncertain estimated value of -1.3 for b (Table V.). For two smaller cycles, with H_{\max} equal to 173 and 57 (*loc. cit.*, figs. 9, 8; Tables VI., V.) the values are -2.0 and -2.4 . For the same material before annealing (*i.e.* in the hard drawn state) the mean values of b'' over the range from I_{\max} to I_{rem} from the results for three cycles, with H_{\max} equal to 351, 175 and 51 (*loc. cit.*, Tables III., II., I.; figs. 5, 4, 3) are -2.9 , -3.3 and -2.6 . These values are numerically greater than those for the annealed material, in contrast to nickel, but in the absence of any satisfactory data for the temperature variation of the magnetostriction of cobalt cannot be usefully discussed at present. In connection with the difference between the mean b'' value, -2.0 , for the annealed material and the estimated value of b , -1.3 , it should perhaps be noted that X-ray examination indicated the presence of a slight amount of cubic phase. Further, the magnetization curve between I_{\max} and I_{rem} differs greatly from that for ideal hexagonal polycrystalline material with random orientational distribution (fig. 3), although there was no evidence of preferred orientation. Although the interpretation of the main features of the magneto-caloric results for cobalt is perhaps more immediately apparent than for the other materials, the quantitative interpretation of the detailed behaviour of the material leaves many problems for future investigation.

Alloys.

The magneto-caloric behaviour of a considerable number of alloys has been examined by Bates and Harrison (1948; see also Bates and Davis, 1948). In a cycle with a maximum field of about 360, the Q' curve for Monel metal (70 Ni, 30 Cu), as the field is changed from $-H_m$ to $+H_m$, is concave upwards, and the maximum cooling is very large, $-Q'$ exceeding 100,000 (*loc. cit.*, fig. 4). The Curie temperature of this material is not greatly above room temperature (an approximate value is 320°K.), so that, as discussed in § 4, the effect of change of intrinsic magnetization will be large. Even greater cooling effects ($-Q'$ exceeding 300,000) were observed in an alloy W.5 (nickel-silicon, 4 Si) (*loc. cit.*, fig. 5). The same general explanation holds here, and also for the iron-nickel alloys (*loc. cit.*, figs. 12, 16). In view of the absence of data from which precise estimates of dI_0/dT could be made for most of these materials (and also in view of the length of this paper), the many interesting results will not be discussed in detail here; it seems quite clear that the predominant contribution to the magneto-caloric effect in many of these alloys is change in intrinsic magnetization, and that a quantitative interpretation of the main features of the curves could, in principle, be developed along the same lines as for iron, nickel and cobalt.

§ 9. CONCLUDING REMARKS.

The illustrative analyses in the preceding section show that, except in the low field range where the magnetization changes are mainly irreversible, the greater part of the adiabatic temperature change accompanying magnetization can be accounted for by a superposition of two effects. These effects are the reversible heating associated with increase in intrinsic magnetization and the reversible cooling associated with increase of apparent magnetization due to rotation of the domain magnetization vectors in the local anisotropic field. The wide differences in the adiabatic heating curves for different materials, considered in relation to the magnetization curves, arise mainly from differences in the relative magnitude of the two effects. In cobalt the rotational effect is predominant, in alloys with low Curie temperatures the effect of change of intrinsic magnetization. In iron and nickel, over the moderate field range, the two effects are of the same order of magnitude. In those cases in which reasonably reliable data are available for the variation with temperature of the intrinsic magnetization and of the magneto-crystalline anisotropy, the agreement between the indirectly estimated adiabatic temperature changes due to the two processes alone, and the total observed changes for well-annealed materials over the range from maximum to zero field is remarkably close, and can leave little doubt as to the coordinating value of the general method of treatment. The agreement, however, even in the most favourable cases, is by no means exact, and at present little progress can be made in the analysis of the observed effects except over those ranges in which the contribution from irreversible changes is relatively small. It is therefore desirable to indicate how a more complete analysis might be made possible, or, looking at the problem from a slightly different standpoint, how fuller information could be obtained from results on magneto-caloric effects about the elementary processes by which changes in magnetization occur.

There are two main requirements. The first is the development of methods by which the reversible and irreversible contributions to the adiabatic temperature change, and to the change in magnetization may be accurately and unambiguously determined separately for any range of the magnetization curve. The general problem and possible methods have already been sufficiently discussed (§ 7).

The second requirement is the accurate determination of the basic magnetic characteristics of the materials whose magneto-caloric behaviour is under investigation. These are the intrinsic magnetization (sufficiently nearly equal, over the ordinary field range, and at temperatures not too near the Curie point, to the technical saturation magnetization), the leading crystalline anisotropy and magnetostriction coefficients, and the temperature coefficients of these. Even if the single crystal behaviour is adequately covered by single anisotropy and magnetostriction coefficients, six quantities are required, namely I_0 , K , and λ , and dI_0/dT , dK/dT and $d\lambda/dT$. It is surprising to find that, for the materials which

have been most extensively studied, iron, nickel and cobalt, it is only for nickel that reasonably reliable estimates of all six quantities at room temperature can be made from published data. The necessary measurements are by no means easy to make, for, apart from the usual difficulties in making any magnetic measurements of high accuracy, unambiguous results for crystal constants are in general obtainable only if good single crystal specimens can be produced. The wider possibilities of the use of polycrystalline material of controllable and determinable orientational distribution should, however, not be overlooked.

If a separation of the reversible and irreversible effects could be made and the basic magnetic constants of the material under investigation were known, far more information could be derived from a study of the magneto-caloric effects than is at present possible. The general methods of analysis which have been illustrated might, for example, lead to much more definite conclusions about the magnitude of the internal stresses in the particular specimens after different heat treatments, and an extension of the methods might yield valuable information about the characteristics of both the reversible and irreversible boundary movement processes in the lower fields, and also about the formation and dissolution of domain boundaries. Such information, in association with that derivable from an analysis of the magnetization curves themselves, may result in much fuller understanding of the complex phenomena underlying the protean variety of form of the ordinary magnetization curves of ferromagnetic materials.

This paper has been concerned with the analysis and interpretation of the experimental results on the adiabatic temperature changes accompanying magnetization, and in this section attention has been drawn to the further work required if the potential value of those results is to be fully realized. It is, however, fitting to conclude with a tribute to the skill which has been shown in the recent experimental work, still happily in progress, on the accurate measurement of these very small temperature changes, which has opened up a new method of attack on the still numerous unsolved problems of the behaviour of ferromagnetic materials.

We are much indebted to Dr. E. P. Wohlfarth for his lively interest and constructive criticisms throughout the course of this work.

SUMMARY.

A theoretical treatment is presented of experimental results on the adiabatic temperature changes accompanying the magnetization of ferromagnetics in low and moderate fields. The relevant thermodynamic relations are derived, in particular those relating the adiabatic change of temperature with field and the change of magnetization with temperature at constant field. Consideration is given to the various elementary processes by which change of magnetization occurs, namely, change of intrinsic magnetization, rotation of the domain magnetization vector in

the anisotropic crystalline field, in its pure form and as modified by strains, and the reversible and irreversible movement of domain boundaries. The thermal changes associated with these processes are considered and the atomic interpretation of the effects is discussed. The total adiabatic change of temperature, expressed as the "heat developed", Q' , as the field is changed from a particular initial value, may be written in the form

$$Q' = a \int d(IH) + \int b'' H dI.$$

The first term corresponds to change of intrinsic magnetization and the positive coefficient a may be calculated from data on the saturation magnetization and its variation with temperature. The value of b'' is then obtained from the magneto-caloric results, and is compared with the value as estimated for particular processes. Detailed analyses are given for iron, nickel and cobalt. For annealed materials the observed adiabatic temperature changes over most of the range from maximum to zero field agree well with those calculated as arising from change of intrinsic magnetization, and rotation of the domain magnetization vectors in the local anisotropic crystalline fields. The major difference between different materials is due to the difference in the relative magnitude of the two effects, one a heating and the other a cooling. Agreement between the estimated and observed values is in no case exact. This is due partly to the uncertainties in the basic data for the materials from which the estimates are made and in the magneto-caloric results themselves and partly to the difficulty of making quantitative estimates of the effect of internal strains. Until more accurate basic data are available, it is impossible to determine whether effects associated with domain boundary formation and dissolution are present. The analysis cannot be satisfactorily extended to regions in which the contribution from irreversible effects is considerable until an unambiguous separation of the reversible and irreversible contributions is made, methods for which are indicated. It is shown that the magneto-caloric results, if supplemented by other data, can provide valuable information about the elementary processes, including the reversible and irreversible boundary movements, over the whole range of a magnetization curve. At present the treatment provides a qualitative explanation of the general course of the observed adiabatic temperature changes, and a quantitative interpretation of the greater part of the reversible contribution.

REFERENCES.

- ADELSBERGER, U., 1927, *Ann. Phys., Lpz.*, **83**, 184.
 AKULOV, N. S., 1933, *Z. Phys.*, **81**, 790.
 BATES, L. F., and DAVIS, J. H., 1948, *Proc. Phys. Soc.*, **60**, 307.
 BATES, L. F., and EDMONDSON, A. S., 1947, *Proc. Phys. Soc.*, **59**, 329. (See also 1948, *Proc. Phys. Soc.*, **60**, 308.)
 BATES, L. F., and HARRISON, E. G., 1948 *a*, *Proc. Phys. Soc.*, **60**, 213; 1948 *b*, *Ibid.*, **60**, 225.
 BATES, L. F., and HEALEY, D. R., 1943, *Proc. Phys. Soc.*, **55**, 188.
 BATES, J. F., and WESTON, J. C., 1941, *Proc. Phys. Soc.*, **53**, 5.

- BECKER, R., and DÖRING, W., 1939, *Ferromagnetismus* (Berlin: Springer).
 (Photo-lithoprint Reproduction, 1943, Ann Arbor, Michigan: Edwards.)
- BITTER, F., 1937, *Introduction to Ferromagnetism* (New York and London: McGraw-Hill).
- BROOKS, H., 1940, *Phys. Rev.*, **58**, 909.
- CONSTANT, F. W., 1928, *Phys. Rev.*, **32**, 486.
- DÖRING, W., 1936, *Z. Phys.*, **103**, 560.
- ELLWOOD, W. B., 1930, *Phys. Rev.*, **36**, 1066.
- GANS, R., 1932, *Ann. Phys., Lpz.*, **15**, 28.
- GUGGENHEIM, E. A., 1936, *Proc. Roy. Soc. A*, **155**, 49 and 70.
- HARDY, T. C., and QUIMBY, L., 1938, *Phys. Rev.*, **54**, 217.
- HONDA, K., ŌKUBO, J., and HIRONE, T., 1929, *Sci. Rep. Tōhoku Univ.*, **18**, 409.
- KAHAN, T., 1938, *Ann. Phys., Paris*, **9**, 105.
- KERSTEN, M., 1931, *Z. Phys.*, **71**, 553; 1943 a. *Phys. Z.*, **44**, 63; 1943 b. *Grundlagen einer Theorie der ferromagnetischen Hysteresis und der Koerzitivkraft* (Leipzig: Hirzel). (Photo-lithoprint Reproduction, 1946, Ann Arbor, Michigan: Edwards.)
- KIRKHAM, D., 1937, *Phys. Rev.*, **52**, 1162.
- LIVENS, G. H., 1945, *Phil. Mag.* [7], **36**, 1; 1947, *Ibid.*, **38**, 453; 1948, *Proc. Camb. Phil. Soc.*, **44**, 534.
- NÉEL, L., 1946, *Ann. Univ. Grenoble*, **22**, 299; 1947, *C.R. Acad. Sci., Paris*, **224**, 1488 and 1550.
- OKAMURA, T., 1936, *Sci. Rep., Tōhoku Univ.*, **24**, 745.
- POTTER, H. H., 1934, *Proc. Roy. Soc. A*, **146**, 362.
- RHODES, P., 1948, *Proc. Leeds Phil. Soc.*, **5**, 116.
- STONER, E. C., 1935, *Phil. Mag.* [7], **19**, 565; 1936, *Phil. Trans. Roy. Soc. A*, **235**, 165; 1937, *Phil. Mag.* [7], **23**, 833; 1944, *J. Inst. Elect. Engrs.*, **91**, 340; 1948, *Phys. Soc. Rep. Progr. Phys.*, **11**, 43.
- STONER, E. C., and WOHLFARTH, E. P., 1947, *Nature, Lond.*, **160**, 650; 1948, *Phil. Trans. Roy. Soc. A*, **240**, 599.
- TOWNSEND, A., 1935, *Phys. Rev.*, **47**, 306.
- VAN VLECK, J. H., 1937, *Phys. Rev.*, **52**, 1178.
- VONSOVSKY, S. V., 1940, *J. Phys., U.S.S.R.*, **3**, 83.
- WEISS, P., and FORRER, R., 1926, *Ann. Phys., Paris*, **5**, 153; 1929, *Ibid.*, **12**, 279.

XLVI. *Investigation of Soft Radiations by Proportional Counters.*
 III. *The Beta Spectrum of Carbon 14.*

By J. ANGUS, A. L. COCKROFT, and S. C. CURRAN,
 Department of Natural Philosophy, The University of Glasgow*.

[Received January 10, 1949.]

SUMMARY.

The proportional counter method has been applied to a study of the β -spectrum of carbon 14 for energies in the range from about 5 kev. to 157.5 kev. (the measured upper limit). While the spectrum appears to be simple, it is found to show considerable departures from the Fermi distribution expected for an allowed transition; the distribution shows a

* Communicated by the Authors.

pronounced maximum at ~ 40 kev. and falls sharply to nearly zero intensity at lower energies. Considerations of half-life and energy limit, together with the shape of the spectrum in the low energy region, give, in view of the known spin change of unity for the transition, better agreement with the Fermi selection rules than with those of Gamow-Teller.

INTRODUCTION.

THE proportional counter technique described in papers I. and II. (Curran, Angus, and Cockroft 1949 a and b) has been applied to the beta-spectrum of C^{14} . As in the case of tritium (Curran, Angus, and Cockroft 1949 b), the source was introduced in gaseous form, here CO_2 . Some of the advantages of the technique over the spectrographic method have already been discussed.

The application of the proportional counter to the study of the carbon spectrum, known to have an upper energy value of 155 kev. approximately, necessitated several developments in the technique. Firstly, because of the relatively large range associated with this energy (~ 28 mg./cm.² in aluminium), the counter used was considerably larger than those used for tritium, and was pressurized to a total pressure of $5\frac{1}{2}$ atmospheres in order to reduce the range of the beta-particles well below the dimensions of the counter. Secondly, a new method of correction for the distortion arising at the ends of the counter was adopted. The procedure consisted in dividing the central wire of the counter into sections joined together by a glass bead; histograms of the pulse distribution were measured, as previously described, for each end, and the undistorted spectrum was obtained by subtraction.

Measurements of the intensity distribution of the beta-particles of C^{14} have been reported recently by Cook, Langer, and Price (1948), using a large high resolution magnetic spectrometer. Their results indicate an upper energy limit of 156.3 ± 1.0 kev., confirming earlier measurements. The distribution obtained does not agree with the Fermi distribution evaluated for this energy; a rather serious departure arises at energies less than 40 kev. The experimental curve shows a relatively sharp maximum at ~ 30 kev. and falls steeply at lower energies, the intensity at 5 kev. being almost zero. The intensity values measured in this region were found to be affected quite appreciably by the thickness of the β -source and the thickness of the mounting.

The shape of the β -spectrum in the low energy region is of considerable theoretical importance and the proportional counter is particularly suitable for studying spectra in this region. The present work includes an examination of the spectrum down to ~ 4 kev.

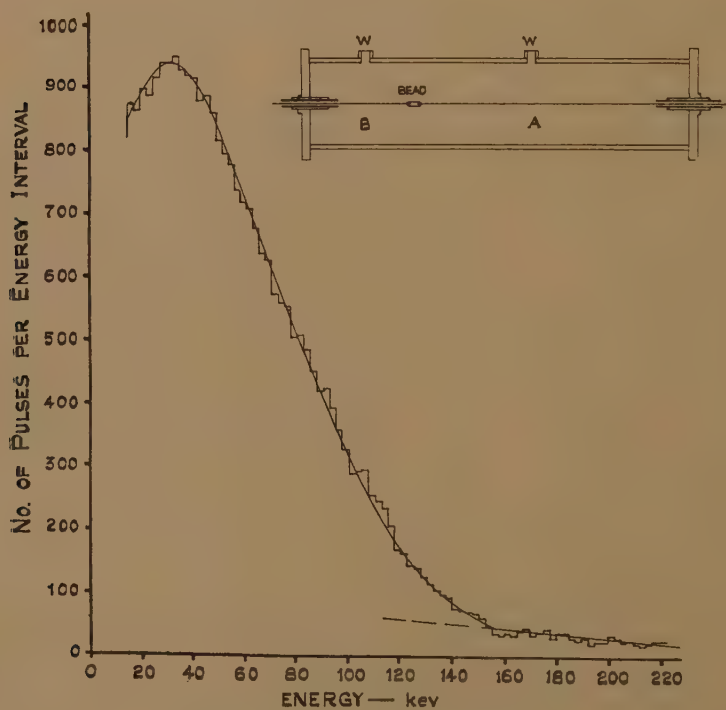
Recent determinations of the half-life of the beta-transition of C^{14} (Hawkins, Hunter *et al.* 1945, Yaffe and Grunland 1945) have given a value between 6 and 7×10^3 years; this result, together with the value of the limiting energy, indicates that the transition is probably second

order forbidden; Konopinski (1943) includes it among second forbidden in his table. The shape of the spectrum for a light element such as C^{14} is therefore of special interest in consideration of such transitions, just as the spectrum of tritium is of particular importance among the allowed transitions.

APPARATUS.

The counter used in this investigation, shown schematically in the inset fig. 1, consisted of a cylinder of copper, 75 cm. long and 14 cm. internal diameter, with thick copper end-plates. The ebonite mountings

Fig. 1.



Intensity distribution in volume A.

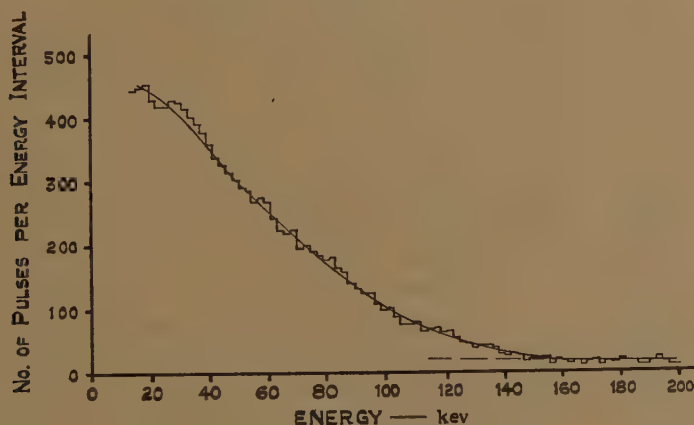
for the central wire and its earthed guard-rings were screwed centrally into the end-plates. The counter had two windows $\frac{1}{4}$ in. diameter covered by an aluminium plate 0.03 in. thick, to allow calibration with low-energy X-rays.

The central wire, tungsten of 0.003 in. diameter, consisted of two portions held together by a little glass rod 1.5 cm. long and 1 mm. in diameter. There was no electrical connection between the two sections and the counter was effectively divided into two independently operating volumes A and B. The lengths of wire exposed in these volumes were 40 cm. and 20 cm. respectively. It was shown in fig. 2 of paper I. that

the gas amplification varied in the vicinity of the end shield and of the bead; each counter section will have both of these variations and each will be of the same magnitude in both sections. If the distribution of pulses is determined at each end at the same gas gain, the distribution obtained by subtracting the histogram for the shorter end from that for the longer end will represent the distribution which would be obtained from a counter of length 20 cm. free from end-effect.

The counter was filled with a mixture of nitrogen at a pressure of 30 cm. Hg and argon at rather more than 5 atmospheres. This reduces the range of a β -particle with maximum energy to $0.4 \times$ radius of the counter. Nitrogen was used since the electron attachment coefficient is negligible, and since the energy required to produce an ion pair has been found to be remarkably constant (Curran *et al.* 1949 a). The voltage required to operate the counter was between 5500 and 6000 volts, and the insulation of the ebonite mountings was very satisfactory. The number of spurious pulses was negligibly small.

Fig. 2.



Intensity distribution in volume B.

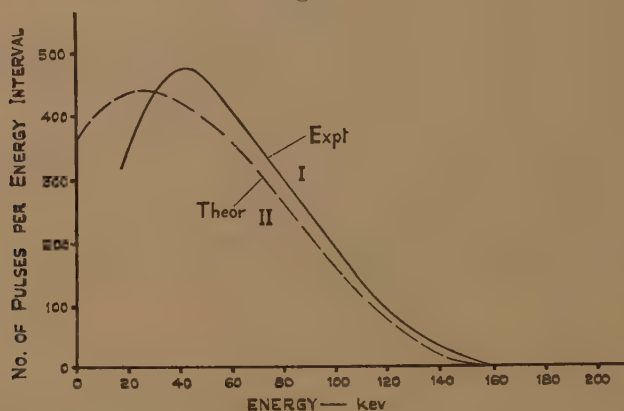
It may be stated here that the radioactive gas used, namely CO_2 , did not interact in any way with the walls of the vessel, since all activity could be removed very easily by evacuating the counter.

PULSE DISTRIBUTIONS.

The radioactive carbon dioxide used in these experiments was prepared by heating a small amount of barium carbonate containing 3 microcuries of C^{14} on a filament sealed into a vacuum vessel containing argon at a pressure of 30 cm. Hg. The argon acted as a carrier for the CO_2 released, and a small fraction of the mixture, giving some 40,000 counts per minute, was introduced into the proportional counter. The counting rates at the two ends were 20,000 per minute and 10,000 per minute, compared with background counting rates of 3000 and 1500 per minute respectively.

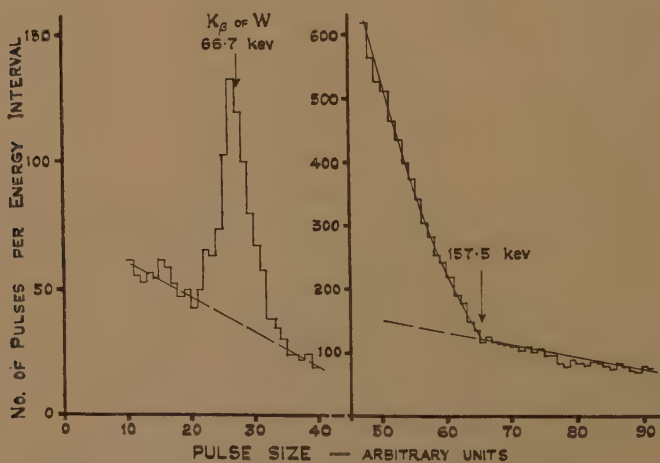
The first set of readings was obtained with a gas gain of about 25 and with the amplifier gain reduced to allow pulses in the energy range from 20 to 200 kev. to be measured. The pulse distribution of the larger volume is given in fig. 1 and that of the smaller in fig. 2: background distributions are not given; they average about 50 pulses per group at

Fig. 3.



Comparison of Fermi allowed distribution with experimental observations.

Fig. 4.



Distribution near end-point with X-ray calibration.

the longer end and 25 per group at the shorter end. Usually the energy scale was set up by superimposing on the spectrum the photoelectrons produced in the counter by the fluorescence X-rays of tungsten. These X-rays were excited in the anode of the X-ray set and were fired directly through the copper wall of the counter ($\frac{3}{8}$ in. thick). A typical calibration histogram is shown on the left-hand side of fig. 4; in each case a single group of pulses was observed, and this group was identified as the K_{β}

group of tungsten at 66.7 kev. The K_{α} group does not appear in appreciable intensity as it is absorbed much more strongly in traversing the copper wall, due to its slightly longer wavelength. In the case of the distribution shown in fig. 1 the energy scale was set up by means of the evaluation of the end point shown in fig. 4.

Fig. 4 gives the pulse distribution in more detail in the energy region between 120 kev. and 200 kev. which was obtained in the longer section and includes the background. The energy value of the end-point is evaluated as 157.5 ± 5 kev.

The smooth curves drawn through figs. 1 and 2 were subtracted and the background difference allowed for: the resulting distribution is given in curve I. of fig. 3. Curve II. of this figure is a plot of the Fermi distribution for an end-point energy of 157 kev. and a nuclear charge of 7,

$$N(W) = p \cdot W(W_0 - W)^2 \cdot G \cdot D,$$

where

p is the momentum in units of mc ,

W is the total energy in units of mc^2 ,

G is the Coulomb factor $\xi \cdot e^{\xi} / (e^{\xi} - 1)$,

$\xi = 2\pi Z / 137 \cdot (W/p)$,

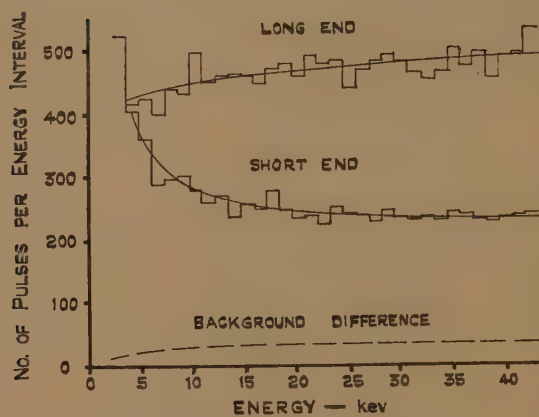
and

D is the Dirac factor.

We have compared these curves with that obtained by Cook, Langer, and Price (1948) (for the thinnest source), correcting for the different abscissæ. We find that whereas the theoretical Fermi distribution has a maximum at about 24 kev., our measured value is 40 kev., while Cook *et al.* find that it occurs just above 30 kev. This difference in the positions of the measured maxima is certainly significant, and the results which we obtained at increased amplifier gain, presented in fig. 6 below, confirm our larger value. We believe that the maximum occurs close to 40 kev. since the errors introduced in our technique tend to depress the position of the maximum to lower values. We consider that our rate of fall below 40 kev. as given in fig. 3 is somewhat exaggerated and that the curve shown in fig. 6 is a truer representation of the spectrum below the maximum.

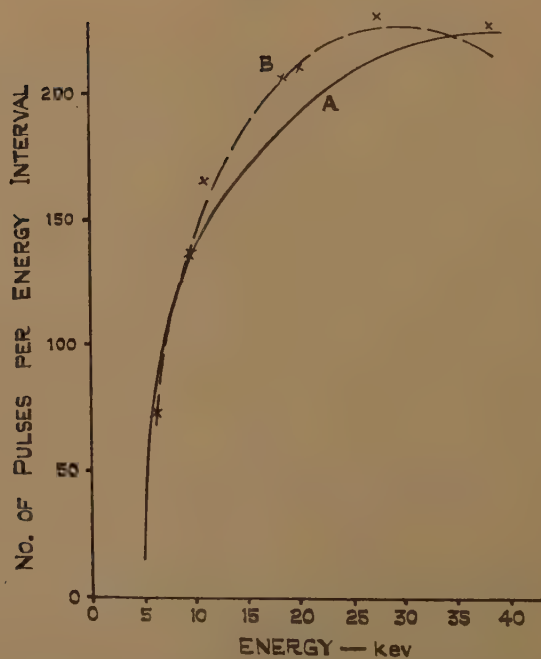
The results of the second investigation, designed to obtain the distribution more accurately between 5 kev. and 40 kev., are shown in figs. 5 and 6. In fig. 5 the pulse distributions obtained at each end are plotted, together with the smoothed background difference between the two ends. Energy scales were set up by superimposing on the particles from the source photo-electrons from the fluorescence X-rays of silver introduced through the windows; K_{α} and K_{β} peaks similar to those of fig. 4 of paper I. were obtained. The difference between these two curves, together with background correction, is plotted in curve A of fig. 6. The statistical errors inherent in the curve below 7 kev. are considerable, but there is every indication that the distribution tends to intensity values very close to zero at an energy of a few kev. The reasonable agreement, particularly

Fig. 5.



Intensity distributions at low energies.

Fig. 6.

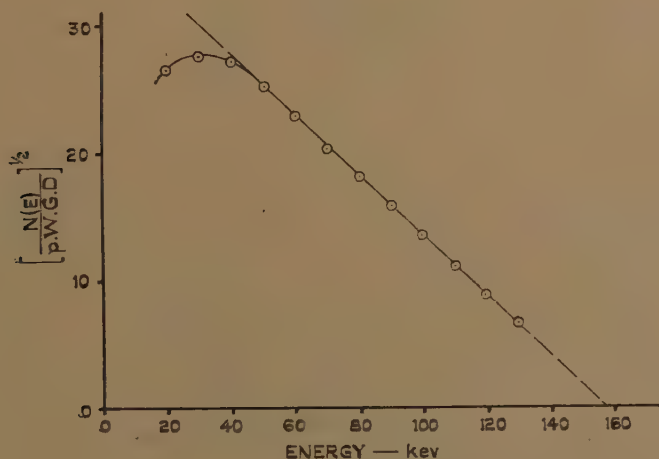


Comparison of present results (curve A) with those of Cook *et al.* (curve B and points X).

at very low energies, between this curve and the results of Cook *et al.*, plotted in curve B, is encouraging, and there can be little doubt that the actual distribution begins to rise rapidly from a very low intensity at a few kev. from the origin, since both experiments find a steep increase. Some theoretical explanation, probably in terms of the "order of forbiddenness" of the transition, is required.

Fig. 7 shows a Fermi plot of the spectrum of fig. 3, *i. e.* $[N(E)/pWGD]^{\frac{1}{2}}$ against E , where $N(E)$ represents the experimental intensity. It will be seen that the experimental curve departs considerably from a linear relation for energies below 50 kev., and also, apparently, for energies between about 130 kev. and the end-point. The latter departure may arise from the small numbers of pulses per group in this region, and it is

Fig. 7.



Fermi plot of experimental results.

not regarded as specially significant, although it would appear that there may be a slight excess of high energy particles. Such an excess could not be explained on the basis of an appreciable neutrino mass, in view of the results for the tritium spectrum (Curran *et al.* 1949 b).

These discrepancies must remain unexplained until a more detailed theory of beta-decay is evolved. The experiment, however, seems to give rather definite support to the Fermi selection rules as opposed to those of Gamow and Teller. The energy limit and life of C^{14} classify it definitely as a second order forbidden transition, since the ft value (Konopinski 1943) is 7×10^8 . Moreover, the general shape of the measured spectrum, suggesting a rather steep parabolic increase near zero energy, seems to support the classification as second or higher order. The spin change involved is unity. Hence, since Fermi rules allow unit spin change in second order transitions and the Gamow-Teller rules do not, the experiment supports the former.

It is a pleasure to thank Professor P. I. Dee for his valuable advice throughout this work, and Mr. B. Touschek for helpful theoretical discussion.

REFERENCES.

- COOK, C. S., LANGER, L. M., and PRICE, H. C., 1948, *Phys. Rev.*, **74**, 548.
 CURRAN, S. C., ANGUS, J., and COCKROFT, A. L., 1949 a, *Phil. Mag.*, **40**, 36-52; 1949 b, *Ibid.*, **40**, 53-60.
 HAWKINS, H. C., HUNTER, R. F., *et al.*, 1948, *Phys. Rev.*, **74**, 696.
 KONOPINSKI, E. J., 1943, *Rev. Mod. Phys.*, **15**, 209.
 YAFFE, L., and GRUNLUND, J. M., 1945, *Phys. Rev.*, **74**, 696.

XLVII. *Origin of Cosmic Ray Stars.*

By J. B. HARDING,
 Imperial College, London *.

[Received January 21, 1949.]

§ 1. INTRODUCTION.

THE nature of the radiation responsible for the production of cosmic ray stars, as observed in photographic emulsions, has been discussed by Perkins (1947 a). By considering the absorption under various substances of the single tracks, which always accompany and indeed which he showed in number to be proportional to the number of stars, Perkins concluded that a large proportion of these disintegrations were initiated by an uncharged and unstable radiation. It was later pointed out by George (1948) that the results obtained by Perkins for the increase in range by a factor of approximately 3 when these particles were absorbed in lead instead of air, could find explanation in the different atomic weights of the two media. He thus showed experimental data were consistent with the assumption that neutrons are the producing agent for these cosmic ray stars.

In about half the disintegrations observed in photographic plates, in addition to long tracks formed by ionizing particles (protons, α -particles, etc.) a short, very heavily ionized track is observed. Measurements of track lengths of all particles ejected from the stars show that few have ranges between 10 and 20 μ , indicating that these very short tracks are essentially different from the longer ones. It is reasonable to suppose that these are caused by residual nuclei or "recoils", left when the original highly excited nuclei have cooled down by particle emission.

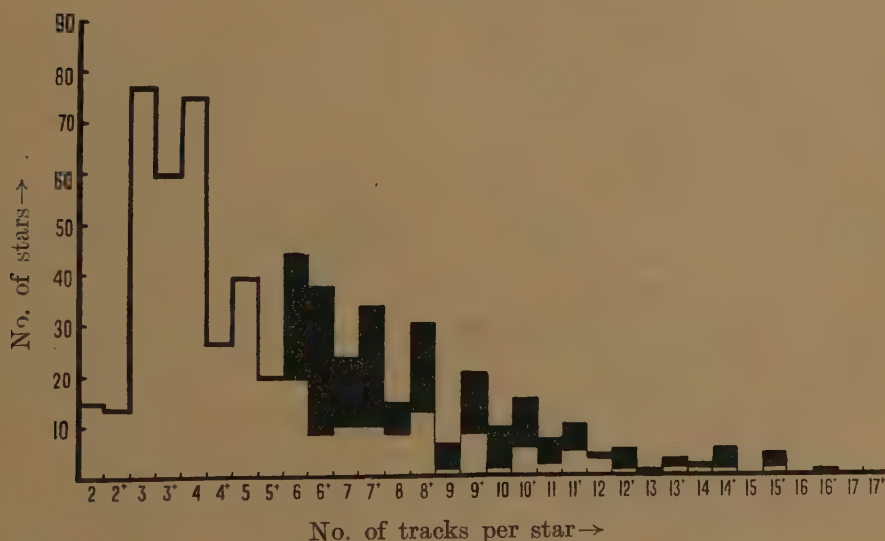
It is from an investigation of the length and angular distribution of these recoil tracks, which are now defined as having ranges less than 10 μ , that further evidence as to the radiation causing nuclear disintegrations can be obtained.

* Communicated by Sir George Thomson.

§ 2. EXPERIMENTAL RESULTS.

The stars investigated were produced in Ilford "Nuclear Research" emulsion coated on glass plates exposed vertically at a height of 3650 m. Two batches of plates were examined. In the first, all the stars (except those produced by σ -mesons) were observed, while in the second set only stars with more than six long tracks were examined. The size distribution of these stars is indicated in fig. 1 in which the superscript "+" after the integer denoting the number of long tracks means that a star with that particular number of prongs also has a visible recoil track with a range of less than $10\ \mu$. For example, in the first batch of plates, twenty stars with six prongs and nine with six prongs and a recoil were observed. In the second batch, the corresponding numbers were twenty-four and twenty-eight, making totals, for both sets, of forty-four six-pronged stars and thirty-seven 6^+ -pronged stars.

Fig. 1.



Size distribution of stars examined.

The unshaded histogram represents the stars examined in the first batch of plates, while the shaded one represents those examined in the second.

The number of α -particles per star was recorded. Phenomenologically, for stars with 2-5 tracks, $110/327=34$ per cent have recoil fragments, and the value of the α/P ratio (total number of α -particles : total number of protons observed in this group of stars) is 0.45. For all the stars with 6-19 tracks, $167/286=58$ per cent have recoils, and the α/P ratio=0.50.

The following measurements were made on each recoil fragment. The length of the projection of the track in the plane of the emulsion and the angle (ψ) made by this projection with the vertical (*i. e.* the azimuthal angle with respect to the vertical edges of the plate) were obtained. The length was estimated to one-quarter of a division of the eyepiece scale

(1 division= $1.4\ \mu$) and the angle to the nearest degree. The depth in the emulsion of the end point of the track relative to the star centre was estimated to one-quarter of a division of the fine adjustment drum on the microscope (1 division $\simeq 3\ \mu$).

If $R = \frac{\text{Number of recoils observed going upwards } (\pi/2 \leq \psi \leq \pi)}{\text{Number of recoils observed going downwards } (0 \leq \psi \leq \pi/2)}$, the

following results were obtained :

FIRST BATCH

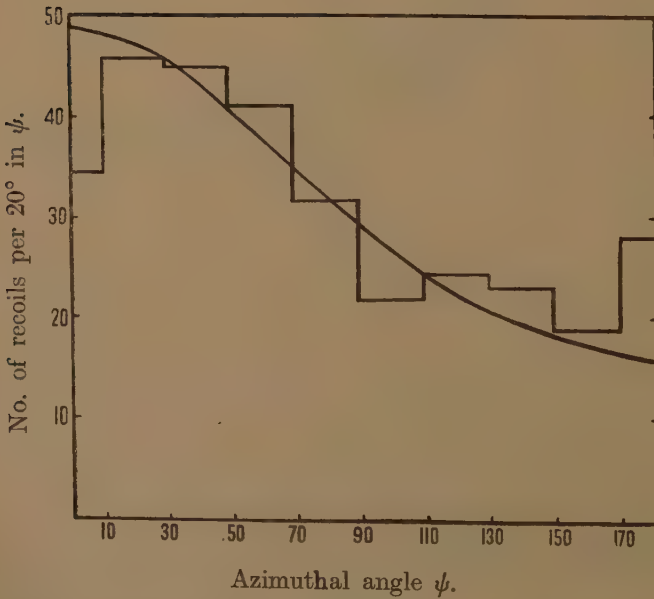
Plate	R
A	$14/31 = 0.45$
B	$20\frac{1}{2}/33\frac{1}{2} = 0.61$
C	$14\frac{1}{2}/20\frac{1}{2} = 0.71$
D	$18/28 = 0.64$

SECOND BATCH

Plate	R
E	$10/13 = 0.77$
F	$8/14 = 0.57$
G	$7/17 = 0.41$
H	$4/9 = 0.44$
I	$6/10 = 0.60$

Total upwards = 102
Total downwards = 177

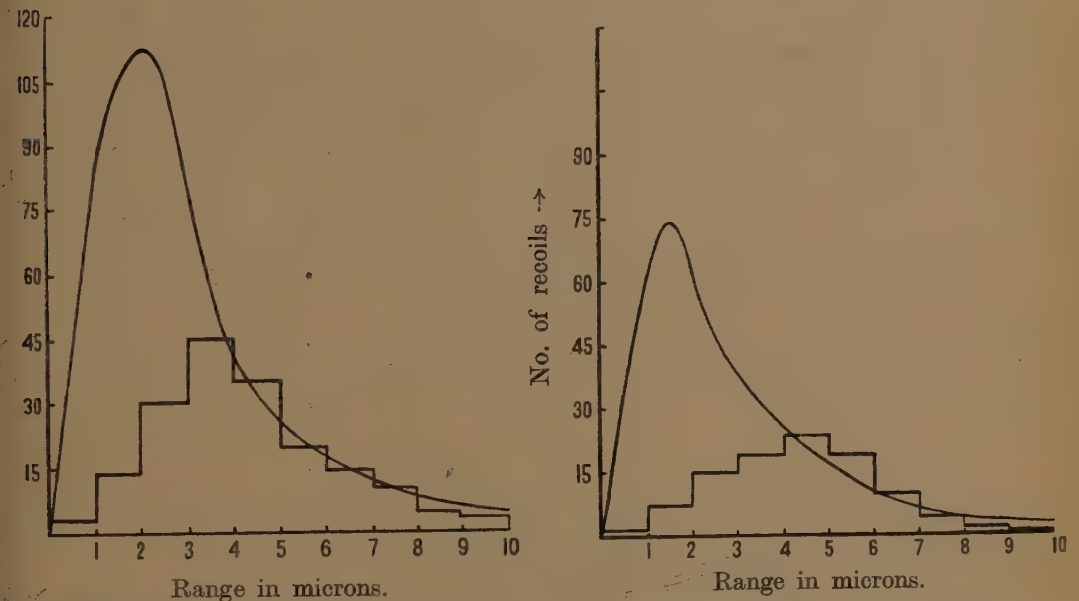
Fig. 2.



Angular distribution of the projection of the recoils. The curve is the theoretical differential angular distribution (see § 3).

The number of recoils per 20° interval in ψ were plotted against ψ , fig. 2, and from the length measurements, the range spectra of the recoils were plotted in fig. 3 (a) for those recoils observed to have travelled downwards (b) for those observed to have travelled in an upward direction.

Figs. 3(a) and 3(b).



(a) Range spectrum of recoil tracks observed to have travelled downwards.

(b) Range spectrum of recoil tracks observed to have travelled upwards.

The theoretical curves have been obtained in (a) by summing curves 2 and 4 in fig. 5, and those in (b) by summing curves 1 and 3 in the same figure.

§ 3. DISCUSSION OF RESULTS.

It is necessary to estimate the mass and charge of a recoil fragment in order to derive any information concerning the nature of the star producing radiation. From unpublished work with emulsions loaded with layers of pure gelatine, the cross-section for star production is seen to be proportional to the geometric cross-section of the nucleus. We can thus define an "average" heavy atom of atomic weight A_0 by

$$\frac{\rho}{A_0} A_0^{\frac{2}{3}} = \sum \frac{\rho_r}{A_r} A_r^{\frac{2}{3}},$$

where ρ = weight in grms. of the heavy atoms (Ag and Br) per cc. of emulsion and ρ_r = weight in grms./cc. of the r th heavy atom of atomic weight A_r . From the constitution of the emulsion $A_0 = 97$, which corresponds to a nuclear charge Ze of $42e$ (where e = electronic charge).

We estimate (see below) that about 500 of the stars examined will originate from heavy nuclei and observe in these that the mean number of visible tracks per star equals six. Using the observed α/P ratio of 0.5, and

assuming equal numbers of protons and neutrons are ejected in a disintegration, we find that the average emitted mass equals $16m_p$ (where m_p =proton mass) and that the emitted charge is $8e$. Therefore we consider a heavy recoil to have a mass $81m_p$ and charge $34e$.

Similar reasoning gives $6m_p$ and $3e$ for the mass and charge respectively of a recoil from an "average" light atom.

A disintegration is produced by some incident particle which, unless it imparts its energy of constitution when at rest, will also give momentum to the nucleus. Suppose the incident particle comes vertically downwards. We can then consider an excited nucleus moving downwards with a velocity V_1 . We have shown (Perkins 1947 b) that such a nucleus cools down by the evaporation of protons, etc. which should be ejected at random with respect to the moving nucleus. In the centre of gravity coordinates, in order to conserve momentum in this process, the residual nucleus, or recoil, must have momentum, P_R , equal and opposite to that of the resultant of all the evaporated particles, P_E . Taken over a sufficiently large number of stars P_E and hence P_R should be isotropic. Therefore in the laboratory system the resultant velocity, V , of an observed recoil is the vector sum of a downward velocity V_1 and a random velocity V_2 where

$$m_R V_2 = P_R = -P_E,$$

m_R being the mass of the recoil fragment.

With the notation as in fig. 4, all recoils ejected within $\theta_0 > \theta > -\pi/2$, where θ is measured with respect to the centre of gravity system, will appear to be going downwards in the plate system.

If $P(\theta) d\theta$ is the probability of ejection between θ and $\theta + d\theta$ we assume $P(\theta) d\theta = \frac{1}{2} \cos \theta d\theta$ (*i. e.* V_2 is a random vector).

Therefore in the plate system

$$\begin{aligned} R &= \frac{\text{No. of recoils observed going upwards}}{\text{No. of recoils observed going downwards}} \\ &= \frac{\frac{1}{2} \int_{\theta_0}^{\pi/2} \cos \theta d\theta}{\frac{1}{2} \int_{-\pi/2}^{\theta_0} \cos \theta d\theta} \\ &= \frac{1 - \sin \theta_0}{1 + \sin \theta_0}. \end{aligned}$$

Experimentally this ratio = 102/177. Therefore $\sin \theta_0 = 75/279 = V_2/V_1$. Therefore $V_2 = 0.27V_1$. If instead of allowing only for vertical incidence of the incoming particles we assume

$$\begin{aligned} P(\gamma) d\gamma &= 3 \cos^2 \gamma \sin \gamma d\gamma \quad (\text{for } 0 < \gamma < \pi/2) \\ &= 0 \text{ otherwise,} \end{aligned}$$

where $P(\gamma) d\gamma$ is the probability of incident particle coming at an angle γ to $\gamma + d\gamma$ with the downward vertical (this relation is indicated by counter

work for the total cosmic ray intensity (Johnson 1933)), it can be shown that

$$V_1 = 4/3 \times 0.27V_2 = 0.37V_2.$$

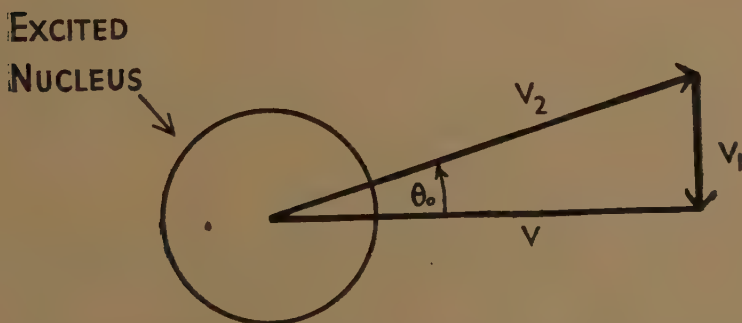
Therefore, taken over all the stars having visible recoil fragments the velocity of the excited nucleus equals 0.37 times the velocity of ejection of the recoil.

To test this hypothesis, we have calculated the differential angular distribution $P(\psi) d\psi$ of the projection of the recoil in the plane of the plate, ψ being the azimuthal angle measured from the vertical ($\psi=0$ is vertically downwards). Using the model outlined above, without, however, considering the angular distribution of the incident particles, we find

$$P(\psi) d\psi =$$

$$N/2\pi \left\{ 1 - \pi/2 (V_1/V_2) \cos \psi + (V_1/V_2)^2 (\cos^2 \psi - \frac{1}{2}) + \text{terms in } (V_1/V_2)^4, \text{ etc.} \right\}.$$

Fig. 4.



V_1 , V_2 are the velocities as previously defined and we have normalized the distribution to the number of observed recoils, *i. e.* $N=279$. This distribution has been plotted in fig. 2 for $V_1 = \frac{1}{3}V_2$, and is in good agreement with the experimental histogram.

By compounding one random vector V_2 with a fixed vector $V_1 \sim \frac{1}{3}V_2$, it can be shown that the average value of the resultant V is about 4 per cent greater than V_2 . Thus by determining the average observed velocity of the recoil fragments the average velocity of the nucleus can be calculated. Experimentally we measure the ranges of the fragments, and as any particular recoil cannot be classified as "heavy" or "light", we assume the majority to be produced from heavy atoms. This assumption is justified below. A range-velocity curve for heavy recoils $A=81$, $Z=34$ is discussed in the appendix. Meanwhile, it may be stated that for an average range of 4.3μ the mean velocity V of ejection of a heavy recoil fragment equals 3.3×10^8 cm./sec.

$$\text{Therefore } V_2 = 3.2 \times 10^8 \text{ cm./sec.}$$

$$\text{Therefore } V_1 = 0.37 \times 3.2 \times 10^8 = 1.2 \times 10^8 \text{ cm./sec.}$$

It should here be noted that this velocity is small compared with the velocity of ejection of the protons, etc. ~ 0.1 times velocity of light. Therefore, if the star particles are ejected isotropically with respect to the moving nucleus, as supposed, they should be practically isotropic in the plate coordinates. This is what we observe in all but the very large stars (>15 tracks).

§ 4. COMPARISON OF RANGE OF RECOILS WITH THEORETICAL RANGE.

We have considered the resultant velocity V of ejection of recoils to be the vector sum of a velocity V_1 and a random velocity V_2 . In practice these velocities will have distributions and not unique values. Assuming that the star particles evaporate, the distribution $\rho(P_E) dP_E$ of their resultant momentum, P_E , can be determined from random walk theory, and hence the differential velocity distribution $\rho(V_2) dV_2$ can be obtained. We will now calculate this distribution, both for "light" and "heavy" fragments, and use it to obtain theoretical distributions for the resultant velocity V .

Experiments with the above mentioned "gelatine" plates show that in a random sample of stars, one-quarter originate in the light elements (C, O, N). We therefore estimate that ~ 110 stars in the first batch are due to light elements, and so for both batches together we are observing 503 disintegration of silver and bromine out of the total of 613 stars. This assumes that the light elements do not produce stars with more than five tracks as indicated by the 30 "gelatine stars" observed in this laboratory. These thirty stars seem to have a similar size distribution to the stars with less than six tracks observed in normal emulsion. We can therefore assign a distribution in the number of visible prongs per star both for heavy and light groups. From the observed α/P ratios we then calculate the distributions $(M_N)_{\text{Heavy}}$ and $(M_N)_{\text{Light}}$ of the *total* number of ejected particles per star, allowing equal numbers of neutrons and protons, *i. e.* M_N is the number of stars with N ejected particles.

We have shown (Perkins 1947 b) that the energy of emission of a star particle is about 10 MeV. From the α/P ratio, the average mass ejected per particle, including neutrons, is $1.6m_p$. Therefore the average momentum of an ejected particle is $\sqrt{(2 \times (1.6 \times 931) \times 10)}$ which equals 175 MeV/c $=P$ (say). It has been shown by Chandrasekhar (1943) that the probability $W(P_E) dP_E$ that the resultant of N random vectors of magnitude P has a magnitude P_E is given by

$$W(P_E) dP_E = (4\pi P_E^2 / (2\pi N P^2 / 3)^{3/2}) \exp \{-(3 |P_E|^2 / 2NP^2)\} dP_E$$

for $N \gg 1$. Exact forms of $W(P_E)$ are given for $N=3$ and 4. The above formula is also true for a Gaussian distribution in magnitude of the vector P if the average value of its magnitude is used in the above equation. Summing such distributions over stars of different sizes, *i. e.* over N , we get for the 503 heavy stars

$$\rho(P_E) dP_E = 4\pi P_E^2 \left\{ \sum_N \frac{\exp \{-(3 |P_E|^2 / 2NP^2)\}}{(2\pi N P^2 / 3)^{3/2}} (M_N)_{\text{Heavy}} \right\}.$$

Since in any particular star $m_R V_2 = P_E$, we can thus obtain the distribution $\rho(V_2) dV_2$, that is, the number of recoils ejected with a velocity between V_2 and $V_2 + dV_2$ in the centre of gravity system, by taking a constant mass of $81m_p$ for heavy recoils.

If θ' is the observed angle in the plate system between the recoil track and the vertical, it may be shown that

$$N(\theta', V) d\theta' dV = \frac{V^2 \sin \theta' \rho(V_1^2 + V^2 + 2V_1 V \cos \theta')^{\frac{1}{2}}}{2(V_1^2 + V^2 + 2V_1 V \cos \theta')} d\theta' dV, \quad (1)$$

where $N(\theta', V) d\theta' dV$ is the number of recoils with resultant velocity V at an angle θ' to $\theta' + d\theta'$.

We have the relation

$$V_2^2 = V_1^2 + V^2 + 2V_1 V \cos \theta'. \quad (2)$$

We thus have

$$\begin{aligned} R &= \frac{\int_0^\infty \int_0^{\pi/2} N(\theta', V) d\theta' dV}{\int_0^\infty \int_{\pi/2}^\pi N(\theta', V) d\theta' dV} \\ &= \frac{\int_{V_1}^\infty (1 - V_1/V_2) \rho(V_2) dV_2}{\int_{V_1}^\infty (1 + V_1/V_2) \rho(V_2) dV_2}. \end{aligned}$$

Equating R to the experimentally observed value and having calculated the function $\rho(V_2) dV_2$ we have an equation for determining V_1 , interpreted as the average velocity of the nucleus. For the heavy stars this equation, when solved numerically, gives $V_1 = 0.43 \times 10^8$ cm./sec. From the observed recoils we estimated $V_1 \sim 1.2 \times 10^8$ cm./sec. The difference in these values would indicate that we are not observing recoils of small velocities, *i. e.* of short range. We do in fact find that only about 50 per cent of the stars have observable recoils.

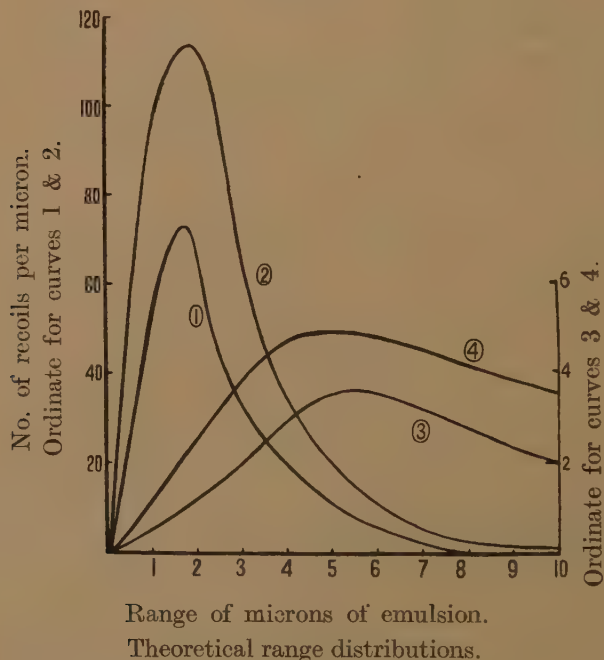
Taking the calculated value $V_1 = 0.43 \times 10^8$ cm./sec., which is obtained above by assuming, for mathematical convenience, that even those recoils overlooked are collimated downwards as are those observed, and using equations (1) and (2) we get

$$\begin{aligned} N(V) dV \text{ (observed upwards)} &= \int_0^{\pi/2} N(\theta' V) dV d\theta' \\ &= -\frac{V dV}{2V_1} \int_{\sqrt{V^2 + V_1^2}}^{V + V_1} \frac{\rho(V_2) dV_2}{V_2}, \\ N(V) dV \text{ (observed downwards)} &= \int_{\pi/2}^\pi N(\theta' V) dV d\theta' \\ &= -\frac{V dV}{2V_1} \int_{V - V_1}^{\sqrt{V^2 + V_1^2}} \frac{\rho(V_2) dV_2}{V_2}. \end{aligned}$$

The integrals for the velocity distribution of recoils from the 503 heavy stars were performed numerically.

From the range-velocity curve discussed in the appendix, and from the calculated functions $N(V)dV$ for heavy recoils, the range distributions for recoils which should travel upwards and downwards were obtained. As we were unable to detect any significant variation in R with star size, a similar analysis was performed for the light recoils. This gave $V_1 = 2.8 \times 10^8$ cm./sec. and with this value the functions $N(V)dV$ (upwards and downwards) were calculated for the light recoils.

Fig. 5.



- (1) For heavy recoils travelling upwards.
- (2) For heavy recoils travelling downwards.
- (3) For light recoils travelling upwards.
- (4) For light recoils travelling downwards.

For these recoils we have

$$R = (m/Z^2 e^2) f(E/m),$$

where R = range of the recoil, m , Ze , E being its mass, charge and energy respectively. The range-velocity curve can be obtained from the corresponding curve for α -particles. Thus range distributions, upwards and downwards, for light fragments were obtained. These four calculated range distributions are shown in fig. 5.

Two points of interest now arise. We see firstly from fig. 5 that although about 110 stars out of the 613 examined will arise through the disintegration of light atoms, only about half of these stars should have recoils

with ranges less than 10μ and thus qualify to be included in our study. In other words, the great majority of recoils observed are recoils from heavy elements. We are thus justified in taking for the average nuclear velocity that velocity V_1 , obtained in the calculations relating to the heavy recoils.

Secondly, it is seen from fig. 3 that the observed and calculated numbers are in good agreement for the longer recoils, but as already suggested, it would seem that we did not observe the shorter ones. We therefore think that the value for the nuclear velocity calculated from random walk theory is a better estimate than that obtained from considerations of the average range of the observed recoils.

We will therefore take a value of $4/3 \times 0.43 \times 10^8$ cm./sec. for the average velocity of an excited nucleus, the factor $4/3$ being now introduced to allow for an angular distribution of the incident particles. (A mathematical analysis on the above lines became too complex if a distribution in angle of the incident particle was allowed from the start.) The standard error on this velocity is difficult to ascertain. We know the mean proton and alpha particle energies to ± 10 per cent and their momenta to ± 5 per cent. Assuming the mean momentum of the neutrons is the same as the protons to ± 50 per cent, the resultant momentum and hence the velocity V_2 is correct to ± 30 per cent.

§ 5. CONCLUSIONS.

The mean momentum transfer from the incident particles to a heavy nucleus is $4/3 \times (0.43 \times 10^8/3 \times 10^{10}) \times 97 \times 931 \doteq 175$ MeV/c. Similarly the momentum transfer to a light nucleus is ~ 160 MeV./c. From (a) the average excitation energy of a star with N tracks, (b) the spread of this energy about the mean value, (c) the size distributions $(M_N)_{\text{Light}}$ and $(M_N)_{\text{Heavy}}$, we have calculated the distribution of excitation energy, that is, the number of stars with excitation energy U to $U+dU$ for the observed groups of heavy and light stars. The results are shown in fig. 6, from which it is seen that the most probable excitation energy for the heavy stars ~ 100 MeV. and that the mean is ~ 190 MeV. For the light stars the corresponding energies are ~ 100 and ~ 130 MeV. *The energy transfer is thus approximately equal to c times the momentum transfer, where c=velocity of light.*

We are now able to discuss the possible star producing radiations.

If the star excitation energy U is derived from the kinetic energy of a neutron which is brought to rest in the nucleus, we have the following formula for the momentum transfer, p ,

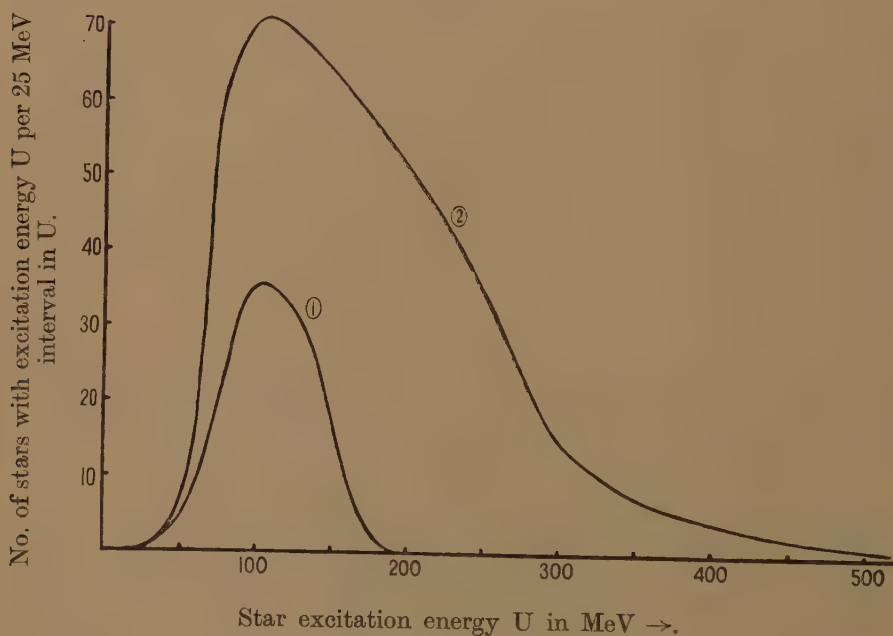
$$pc = \sqrt{(2m_0c^2U + U^2)},$$

where m_0 is the rest mass of the neutron. We see that for an energy transfer ~ 190 MeV. the corresponding momentum transfer is ~ 600 MeV./c. We can thus discount captured neutrons as being the radiation responsible for the majority of stars.

A neutron which excites a nucleus without being captured would lose momentum in excess of $1/c$ times the energy transfer : in the limit, when moving with the velocity of light, the energy transfer equals c times the momentum transfer. The results obtained are therefore not inconsistent with the view that the majority of stars are produced by very fast neutrons.

Although the required equality in the energy and momentum transfer is obviously satisfied in the case of photons, it would seem that they are not responsible for more than a few per cent of the stars as, in cloud chamber work, such disintegrations are rarely observed in connection with cascade showers.

Fig. 6.



Distribution of star excitation energy.

(1) For the 110 "light" stars.

(2) For the 503 "heavy" stars.

If we now assume that stars are produced by unstable particles, two cases arise. Firstly, the particles may give up to the nucleus all their rest mass, plus any kinetic energy. Therefore in the equation

$$m_0c^2 = \sqrt{(E^2 - p^2c^2)}, \quad \dots \dots \dots (3)$$

where E is the total energy of a particle, we associate E with the excitation energy of the star, U . Taking values 190 MeV. and 175 MeV./ c . for E and p , equation (3) is solved for $m_0c^2 = 74$ MeV. We see then that the results can be explained if mesons of mass $\sim 150m_e$ (m_e =electron mass) are captured while moving with relativistic velocities.

Secondly, a meson may give up only part of its rest energy to the star. For example, a meson of mass $\sim 500 \text{ MeV./c}^2$, if such exist, might give up its total energy to one or more nucleons. These nucleons, escaping from the nucleus may lose only part of their energy and it is conceivable, that if these mesons come predominantly downwards, and the interaction occurs in the upper part of the nucleus, the momentum transfer will also be mainly downwards. The required energy and momentum transfer could be obtained on this model.

The possible processes indicated above are essentially alternative processes, inasmuch as any one producing all, or at least the majority of stars, would satisfy the conditions of momentum and energy transfer. As yet we have not ruled out the possibility that perhaps half the stars are produced in a process which causes a large energy transference for only a small transfer of momentum, while in the other half, the momentum transfer is much greater than the corresponding energy transfer. Such would be the case, for example, if half the stars were produced by the decay, when at rest, of neutral mesons, the excitation energy for the others coming from the kinetic energy of captured neutrons.

If however, the majority of stars were produced by two such essentially different processes, we should expect to observe more recoils, by a factor of about three, with ranges between 5 and 10μ , travelling downwards, and fewer travelling upwards. This hypothesis is thus discredited. The good agreement in this range between the observed and calculated distributions (figs. 3(a) and 3(b)) supports the hypothesis that the majority of stars are produced by one type of process.

It therefore seems that, either fast neutrons which are not captured, fast mesons which give up all their rest energy, or possibly "heavy mesons" which are captured in the upper part of nuclei, are responsible for the majority of cosmic ray stars observed in photographic emulsions.

Note added in proof.—Further work with "gelatine" plates (see *Nature, Lond.*, in the press) indicates that the cross-section for disintegration of the light nuclei relative to the heavy, is greater than the ratios of their geometrical cross-sections. In a random sample of stars in normal emulsion 36 ± 7 per cent of the stars are now due to disintegrations of light nuclei. Therefore in the two batches of stars examined, 150 are due to light nuclei, instead of 110 as used in the text, and 453 instead of 503 are due to heavy nuclei. In fig. 5 the ordinate for curves 1 and 2 should now be multiplied by $\frac{453}{503}$ and for curves 3 and 4 by $\frac{150}{110}$. This will improve the agreement between the observed and calculated range distributions shown in fig. 3. The curves of fig. 6 should also be slightly modified but the conclusions drawn therefrom remain unaltered.

ACKNOWLEDGMENTS.

I wish to express my appreciation to Professor Sir George Thomson for his continued help and advice during the course of this work, and also to thank Mr. N. C. Barford who derived a mathematical expression used in this paper.

APPENDIX.

RANGE-VELOCITY RELATION FOR RECOIL FRAGMENTS IN PHOTOGRAPHIC EMULSIONS.

It has been suggested by Bohr (1941) that, when considering the energy loss of a highly charged fragment, two stopping mechanisms have to be taken into account. When the fragment velocity is high, its effective charge is large and the stopping is caused mainly by energy transfer to the individual electrons of the medium (electronic stopping). With decrease of velocity, and hence also of effective charge, energy loss through direct momentum transfer from the fragment to the nuclei of the medium will become of greater importance (nuclear stopping).

Experimental range-velocity curves have been obtained for uranium fission fragments in argon gas by Bohr *et al.* (1940) (fig. 7). From these curves, by employing theoretical formulæ derived by Bohr (1941), we have (a) to deduce a range-velocity relation for a fragment of different mass and charge (a typical star recoil has $A=81$ and $Ze=34e$) in argon and then (b) recalculate the relation for a medium of photographic emulsion.

Bohr derived the following formula for the rate of velocity loss of a fragment, mass M_1 charge Z_1e :

$$\frac{1}{N} \frac{dV}{dx} = \frac{4\pi e^4}{M_1 m V^3} \left(Z_1^{\frac{1}{2}} \frac{V}{V_0} \right)^2 \sum_s \ln \frac{m V^2 V_0}{2\pi v_s e^2 Z_1^{\frac{1}{2}}} + \frac{4\pi e^4 Z_1^2 Z_2^2}{M_1 M_2 V^3} \ln \frac{M_1 M_2 V^2 a_{12}}{(M_1 + M_2) Z_1 Z_2 e^2}, \quad \dots \dots (1)$$

where N is the number of gas atoms per unit volume, e and m are the electronic mass and charge, Z_2e and M_2 the charge and mass of the gas nuclei

$$a_{12} = a_0 (Z_1^{\frac{2}{3}} + Z_2^{\frac{2}{3}})^{-\frac{1}{2}},$$

where a_0 = radius of the hydrogen orbit, V = velocity of the fragment, V_0 = velocity of the electron in the hydrogen atom $\sim 2.2 \times 10^8$ cm./sec., v_s = frequency of the s th virtual atomic oscillator. The first term in equation (1) accounts for the contribution to the velocity loss due to electronic collisions and is more important, for $V \gg V_0$, than the second, which accounts for the velocity loss in nuclear collisions. The reverse is true for $V \ll V_0$.

Thus by neglecting the second term, Bohr showed that for any initial velocity V_i , large compared with V_0 the range R_F of a fragment compared with that, R_α , of an α -particle with the same velocity is given by

$$R_F/R_\alpha = 5(M_1/Z_1^{\frac{2}{3}})(V_0/V_i)^2. \quad \dots \dots (2)$$

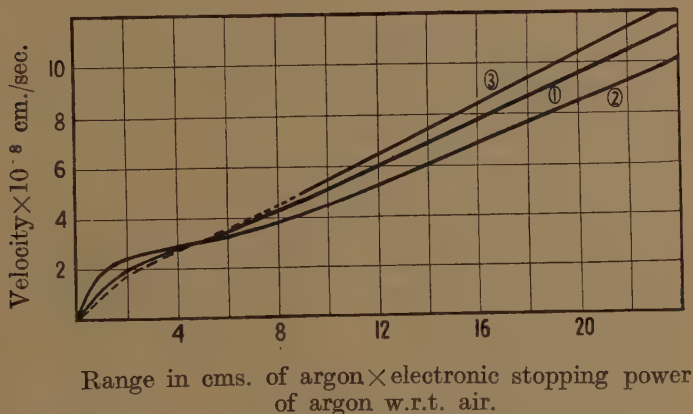
R_F should be interpreted as the range above that obtained by extrapolating the initial linear slope of the curve at high velocities to zero velocity. This ratio is independent of the nature of the gas atoms.

Hence

$$R_F/R'_F = (M_1/M')(Z'_1/Z_1)^{\frac{2}{3}}, \quad . \quad . \quad . \quad (3)$$

where R_F is the range of a second fragment of mass and charge M'_1 and $Z'_1 e$ respectively.

Fig. 7.



1. Experimental range-velocity curve for light uranium fission fragments, $A=95$, $Ze=38$, in argon.
2. Experimental range-velocity curve for heavy uranium fission fragments, $A=139$, $Ze=54$, in argon.
3. Calculated range-velocity curve for recoil fragments $A=81$, $Ze=34$, in argon.

Bohr has shown that, comparing the range R_0 for a fragment of velocity V_0 , deduced from (1) by neglecting the first term, with the range R_F of a fragment with initial velocity V_i , estimated from this term alone, we find

$$R_0/R_F = 0.07 (M_2/m) Z_1^{-\frac{2}{3}} Z_2^{-\frac{2}{3}} (V_0/V_i). \quad . \quad . \quad . \quad (4)$$

Thus for any given media (in this case argon)

$$R'_0/R_0 = R'_F/R_F (Z_1/Z'_1)^{\frac{2}{3}},$$

whence, from equation (3) we get

$$R'_0/R_0 = M'_1/M_1 (Z_1/Z'_1)^2. \quad . \quad . \quad . \quad (5)$$

It can be seen from fig. 7 that for velocities large compared with V_0 , relation (3) is approximately correct for the heavy and light fragments of uranium. Using curve (1) of this figure and equation (3), we calculate

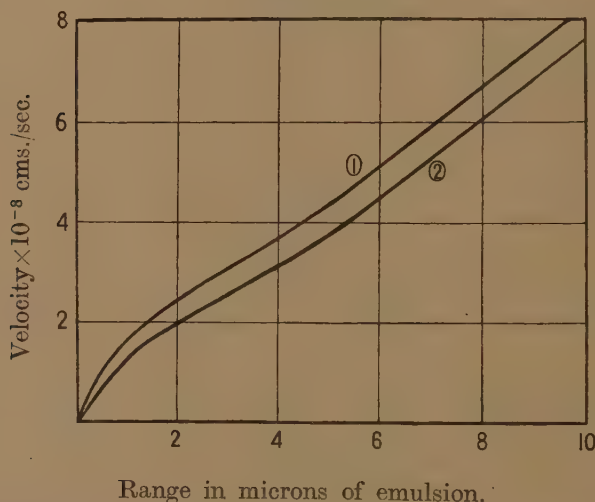
the range-velocity relation for a fragment of mass $81m_p$ and charge $34e$. This curve, 3, was drawn only for velocities above 4×10^8 cm./sec., below which it was estimated by eye. To help this estimation we have calculated, from equation (5), the range R_0 for our fragments in argon, in terms of the experimental range, R'_0 , of the light uranium fission fragment. We have thus obtained a range-velocity curve for recoil fragments in argon gas and it is now necessary to obtain the corresponding relation for such fragments in photographic emulsion.

It follows from (4) that if the suffixes A and E refer to argon and emulsion respectively

$$\frac{(R_0)_A (R_F)_E}{(R_0)_E (R_F)_A} = \frac{M_A Z_A^{-3/2}}{M_E Z_E^{-3/2}}, \quad \dots \dots \dots (6)$$

where M_E and $Z_E e$ are the effective mass and charge of emulsion nuclei.

Fig. 8.



1. Calculated range-velocity curve for recoil fragments in emulsion. Curve not corrected for nuclear stopping.
2. Calculated range-velocity curve for recoil fragments in emulsion. Curve corrected for nuclear stopping and is the relation to be used.

Using equation (2), valid only for large velocities, we get

$$\left(\frac{R_F}{R_\alpha}\right)_E = \left(\frac{R_F}{R_\alpha}\right)_A,$$

$$\therefore \frac{(R_F)_E}{(R_F)_A} = \frac{(R_\alpha)_E}{(R_\alpha)_A} = \frac{\text{electronic stopping power of argon w.r. to air}}{\text{electronic stopping power of emulsion w.r. to air}} \quad \dots \dots \dots (7)$$

Therefore from (6) and (7)

$$\frac{(R_0)_A (R_\alpha)_E}{(R_0)_E (R_\alpha)_A} = \frac{(A/Z^{3/2})_A}{\sum_r \frac{\rho_r}{Z_r^{3/2}} \bigg/ \sum_r \frac{\rho_r}{A_r}},$$

where ρ_r , A_r , Z_r are the density in g./cc., the atomic weight and nuclear charge of the r th component of the emulsion. From the composition of the emulsion we get

$$\begin{aligned} (R_0)_E &= \left\{ \frac{(R_0)_A}{0.69} \right\} \left\{ \frac{(R_\alpha)_E}{(R_\alpha)_A} \right\} \\ &= \frac{(R_0)_A}{0.69} \frac{\text{electronic stopping power of argon w.r. to air}}{\text{electronic stopping power of emulsion w.r. to air}}. \quad (8) \end{aligned}$$

A range energy curve for α -particles in air has been fully discussed by Bethe and Livingston (1937). Experimental curves for these particles in nuclear research emulsions have been obtained by Fowler *et al.* (1947) and so we can obtain the electronic stopping power of emulsion with respect to air. (This is nearly constant in the velocity range considered.) Neglecting for the moment the nuclear stopping, we obtain a velocity-range relation curve (1) fig. 8, for a recoil fragment in photographic emulsion, by dividing the abscissa in fig. 7 by the electronic stopping power of emulsion. In order now to correct for the nuclear stopping we see from equation (8) that the range R_0 , calculated without considering this effect, must now be increased by the factor 1/0.69. The nuclear is more important than the electronic stopping power for all velocities less than V_0 and so all ranges up to R_0 were increased by this factor, thus obtaining the low velocity portion of curve (2) fig. 8. This curve is then continued to higher velocities by drawing it parallel to curve (1) as indicated by equation (7). It is this curve (2), fig. 8, which has been used to represent the range-velocity relation for recoil fragments in photographic plates.

REFERENCES.

- BOHR, N., 1940, *Phys. Rev.*, **58**, 839; 1941, *Ibid.*, **59**, 270.
 CHANDRASEKHAR, S., 1943, *Rev. Mod. Phys.*, **15**, 1.
 FOWLER, P. H. *et al.*, 1947, *Nature, Lond.*, **159**, 301.
 GEORGE, E. P., 1948, *Nature, Lond.*, **162**, 333.
 JOHNSON, T. H., 1933, *Phys. Rev.*, **43**, 703.
 LIVINGSTON, M. S., and BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9**, 262.
 PERKINS, D. H., 1947 a, *Nature, Lond.*, **160**, 707; 1947 b, *Ibid.*, **160**, 299 (see also HARDING *et al.*, *Proc. Roy. Soc.* (in the press.).

XLVIII. *Characteristics of the Crater of the High Intensity Carbon Arc.*

By V. J. FRANCIS, B.Sc., F.Inst.P., M.I.E.E., F. S. HAWKINS, Ph.D.,
A.R.I.C., and A. H. WILLOUGHBY, B.Sc.(Eng.) *.

(Communication from the Staff of the Research Laboratories of The General
Electric Company Limited, Wembley, England.)

[Received January 31st, 1949.]

LIST OF SYMBOLS.

d_1 =diameter of positive (mm.).

d_2 =diameter of core (mm.).

V =voltage across the lamp (volts).

I =current through the lamp (amperes).

P =power in the lamp (kilowatts).

b =burning rate of positive carbon (mm. per hr.).

d_c =crater diameter (mm.).

h_c =crater depth (mm.).

p =positive projection (mm.).

k =rate of decrease of diameter (mm. per hr.) of positive carbon due to external oxidation.

$$K' = \frac{2l_1 p}{d_1} = \frac{K}{d_1}.$$

E =rate of evaporation from crater (mm.³ per hr.).

l_1 =rate of recession of the outer surface normal to itself due to oxidation (mm. per hr.).

l_2 =rate of recession of the inner sloping surface of the crater normal to itself (mm. per hr.).

ψ =angle between positive and negative carbons.

ϕ =angle of slope of external surface of positive carbon (see fig. 5).

θ =angle of slope of the side of the crater (see fig. 5).

m, n , constants (see equation (5, 2.3)).

α, β constants (see equation (7.1)).

SUMMARY.

The paper describes experimental investigations on the burning characteristics of a range of high intensity positive carbons. The properties studied were the rate of burn of the positive carbon with respect to the effects of external oxidation and internal evaporation (§ 5); the crater diameter and depth (§ 6); the arc voltage (§ 7); and the axial candles (§ 8).

* Communicated by the Authors.

There exists, so far as the authors are aware, no theory of operation of the high intensity arc which supplies an explanation, even in a qualitative manner, of the mechanism of the arc in relation to the crater formation and other associated phenomena.

An attempt at a theory is made in this paper, and the experimental results support reasonably well the assumptions on which the theory is based.

An adequate explanation should be capable of predicting quantitative parameters such as the rate of burn and crater diameter, when certain constants of the lamp and carbons and the power input to the lamp are given.

While the present theory does not purport to be complete, there is little doubt that it is basically true and provides a description of the essential mechanism which effectively coordinates the experimental results.

§ 1. INTRODUCTION.

DURING the period 1940-41, in connection with the development of certain special types of high intensity arc, it became necessary to make observations on the behaviour of the positive crater, the rate of burn and the relative axial candles, for a large range of arc currents and for different sizes and types of positive carbon.

It became evident that interesting relations existed between some of the observed quantities, and, since so far as the authors are aware this information is not generally available, it has been thought worth while to put it on record together with some comments (Finkelburg 1940).

A comprehensive mechanism is suggested on the basis of the results obtained, which, if the values of certain parameters were known accurately for any particular combination of carbon and lamp, would enable the entire behaviour of the lamp in respect of the positive crater formation, the electrical operating conditions and the luminous output, to be determined as a function of the power input to the lamp.

All the readings given subsequently were taken when the arc was burning in the quiet and stable state except where otherwise indicated. Few readings only were taken when the arc was in the "exploded" form, under which conditions entirely different relations between the various parameters considered here would exist.*

* At currents in excess of about 450 amperes, the arc undergoes a radical change of shape and becomes noisy. The flame near the positive decreases in height and resembles a bats-wing burner in shape; near the negative it increases in diameter rapidly with increase of distance from the negative tip, and so takes the shape of an inverted truncated cone. The change is believed to arise from a disturbance of the cathode spot, and the new type has been called an exploded form of the normal flame. It has been fully described elsewhere (*Symposium on Searchlight, I.E.S.*, 1948).

§ 2. ARC LAMPS.

Three different lamps, A, B, C, were used for the experiments ; the significant details of their construction are given in Table I.

Two of the lamps, A and B, were designed to take 16 mm. positives only, the third lamp, C, was an experimental lamp constructed for this investigation. It was capable of burning the full range of positives used.

TABLE I.

Lamp	Positive projection (mm.)	Positive rotation	Magnetic control	Angle between positive and negative	Arc length (mm.)
A	35	Yes	Yes	75°	22
B	50	No	Yes	66°	22
C	95	Yes	Yes	45°	28

§ 3. CARBONS.

The range of positive carbons used included those with external diameters varying from 16 mm. to 30 mm. with different core sizes ranging from 7 mm. to 18 mm. All cores had the same chemical composition and consisted of a mixture of 40 parts of amorphous carbon with 60 parts of cerium fluoride ; likewise the shells were all of the same composition, being made from a coke base.

12 mm. negative carbons were burnt with 16 mm. and 20 mm. positives ; 16 mm. negatives were used when 24 mm. and 30 mm. positives were burnt.

§ 4. EXPERIMENTAL.

Observations fall mainly into three categories : those relating to rate of burn of positive carbon and the formation of the positive crater ; those relating to electrical properties ; and observations on the light output.

In measuring the rate of burn of a positive and the dimensions of its crater the procedure was the straightforward one of first burning the arc until stable conditions were reached and then burning at a steady current for a predetermined time (usually five minutes). The positions of both positive and negative carbons were so controlled, partly automatically and partly by manual control, that the arc voltage drop in these circumstances remained constant, so that the power input was also substantially constant. The amount of positive burnt during the period could easily be measured.

The crater depth and diameter were measured after the positive had been allowed to cool. The latter was defined as the diameter of the circle formed by the highest points of the lip. In taking these two measurements only, we had in mind that the crater could, as we subsequently verified, be treated to a fair degree of approximation as a truncated cone with the core itself forming the smaller section. It is important to understand that

the accuracy of these measurements on the crater dimensions is necessarily poor, and the consequences of this are indicated at the appropriate places in the paper.

In determining the relative axial candles, a rectifier type photocell was mounted in a metal box with circular hole cut in its front and covered by a piece of opal glass. By means of a lens an image of the positive crater was projected on to the opal glass, a small metal shield being used to prevent light from the arc flame reaching it.

Results were obtained for lamps A and C, but not for lamp B. The optical system and general arrangement of apparatus were different for the

Fig. 1 a.

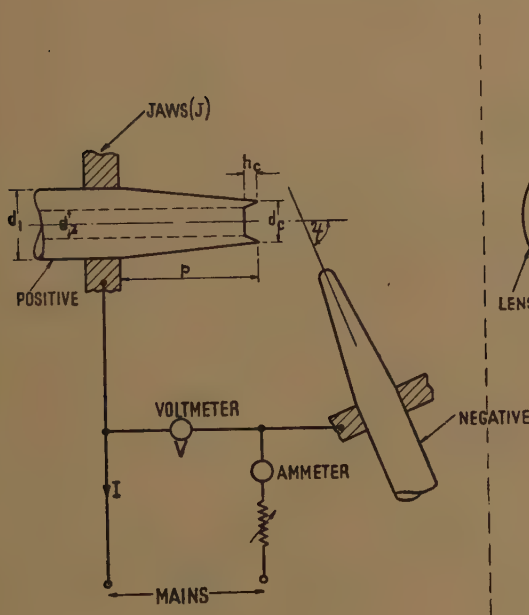
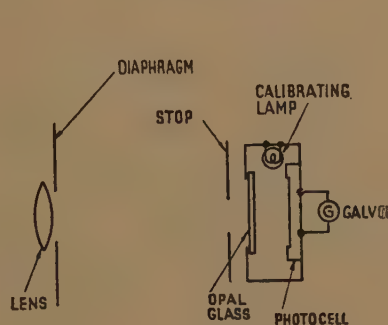


Fig. 1 b.



two lamps, so that since a purely arbitrary scale was used, correlation of the results given by different lamps was not convenient, but, however, was not necessary for our purposes.

On the other hand, a tungsten lamp incorporated in the photocell box enabled frequent calibration of the photocell sensitivity to be made so that readings on any one lamp are strictly comparable. For both lamps the photocell was situated about ten feet from the positive crater; it was connected to a galvanometer, the usual precautions being taken to ensure that the linearity and fatigue properties of the cell did not vitiate the accuracy of the results. Since the cores were all of the same composition, changes in the spectral intensity distribution were not such as to give difficulty in this type of investigation.

Fig. 1 a shows schematically the lamp arrangement, and fig. 1 b the photocell and light-measuring equipment.

§ 5. RATE OF BURN OF POSITIVE.

Experimental Results.

Figs. 2 and 3 show the rate of burn of positive carbons in lamps A and B respectively for 16 mm. carbons with 7.3, 10.2, 11.3 mm. cores, plotted against power input to lamp. It appears that there is a small difference in the behaviour of the two lamps. This may be due to experimental error, although it is known that certain design features of a lamp, such as the distance the positive projects from the jaws through which the current is led, can influence the effective rate of oxidation of the shell. The important conclusion, however, is that the size of the core has no apparent effect on the rate of burn for a given power input.

Fig. 2.

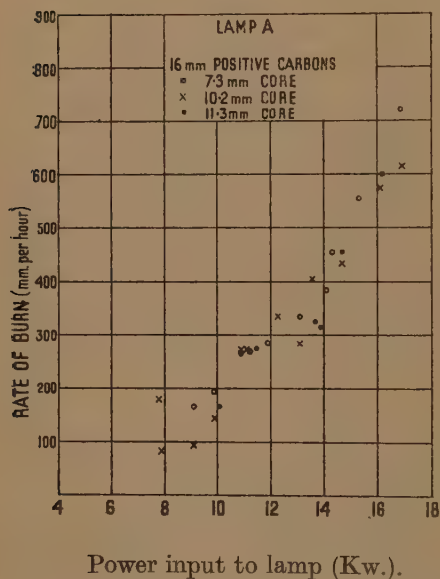
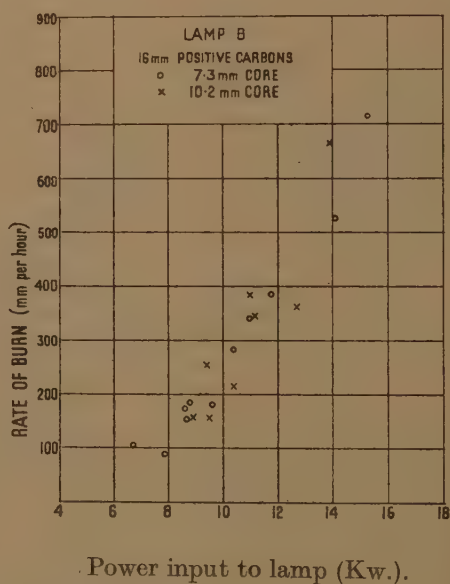


Fig. 3.

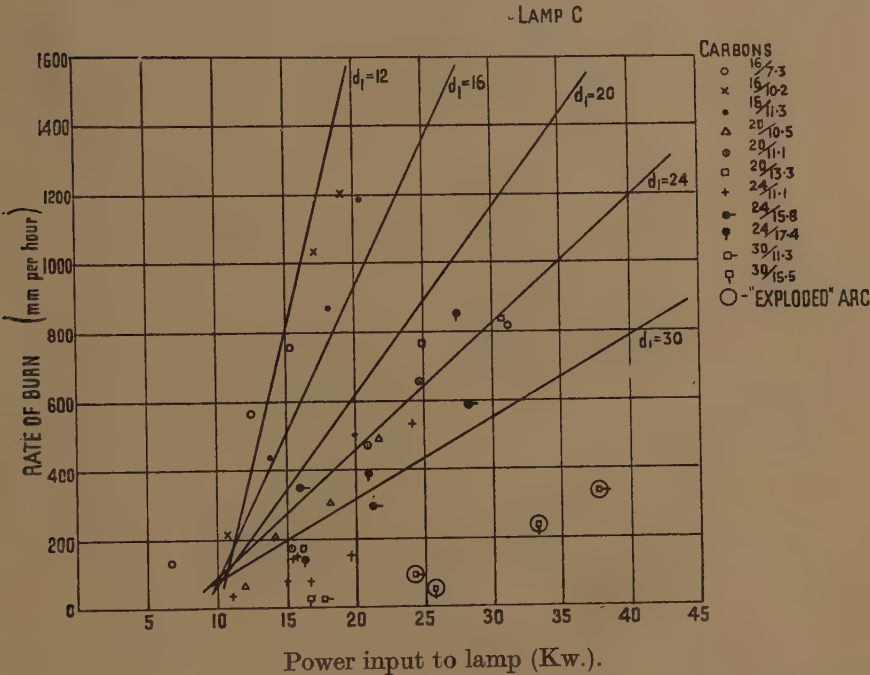


A further conclusion would seem to be that the power dissipated outside the crater is independent of the core size for a given diameter of carbon, as otherwise the rate of burn, which depends mainly on the energy dissipated in the crater, would be different for different core sizes at the same power input. This conclusion is also reached in § 7.

Fig. 4 shows the rate of burn plotted against the power in the arc as a function of the positive electrode used, the points referring to particular combinations of carbon diameter (d_1) and core diameter (d_2). It is clear that for any particular carbon diameter there is a substantially linear relation between the rate of burn and the power input to the lamp, and its slope decreases with increasing carbon size. A similar relationship

exists between the rate of burn and crater diameter. In each case, however, there are four points which do not conform to the above conclusion, and they refer to arcs with exploded flames.

Fig. 4.



Discussion.

1. Oxidation External to Crater.

It was thought worth while to see whether these data provided any information as to the mechanism of the consumption of the positive electrode. In particular, it should be possible to show whether the external surface is merely burnt away by an exothermic oxidation reaction which renders it largely unnecessary for power to be supplied from the arc for this purpose.

If the outer surface is lost by such a process, in any particular lamp the rate of decrease of diameter resulting therefrom will be independent of the size of carbon and of the power input to the arc, although there may be variations from lamp to lamp if the cooling and ventilation conditions differ. The oxidation begins immediately after the positive carbon leaves the positive jaws J (fig. 5) and continues until the crater rim is reached.

If the recession normal to the surface due to oxidation is l_1 mm. per hour, the rate of decrease of diameter is given by

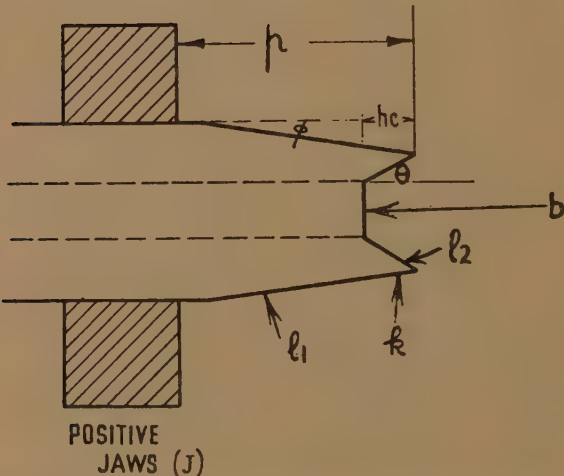
$$k = 2l_1 / \cos \phi, \quad \dots \dots \dots (5, 1.1)$$

where ϕ is the slope of the positive projection (fig. 5). Then, with the burning rate of b mm. per hour, and a positive projection from the jaws of p mm. the time required for a point on the carbon to travel from the positive jaw to the crater is p/b hours and the decrease of diameter is kp/b mm.

Thus

$$\left. \begin{aligned} d_1 - d_c &= kp/b, \\ &= \frac{2l_1 p}{b \cos \phi} \end{aligned} \right\} \dots \dots \dots (5, 1.2).$$

Fig. 5.



For all the measurements with which we are concerned, it may be easily verified that $\cos \phi$ is nearly unity and its variation may be ignored. Then we may write

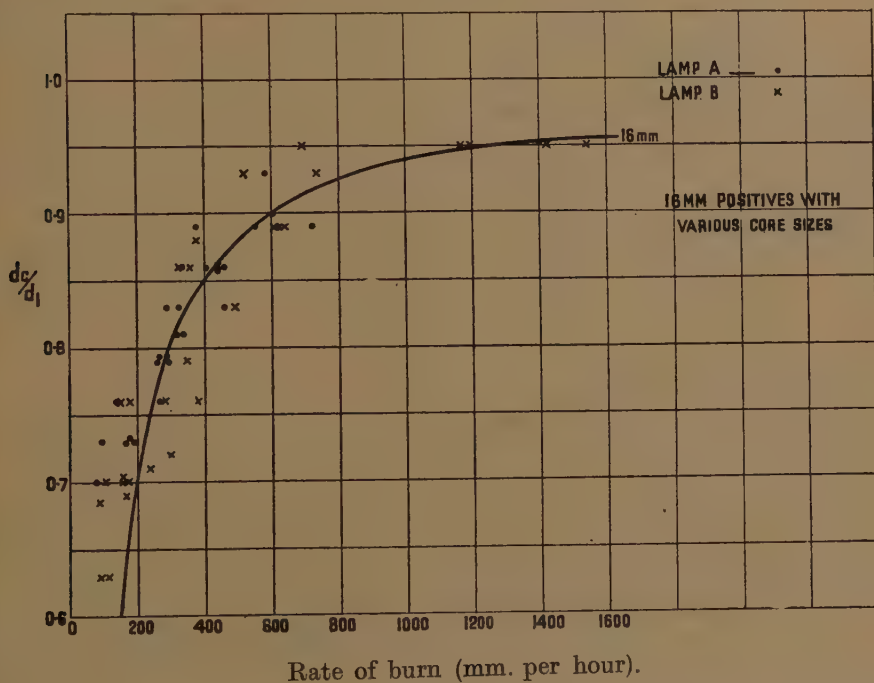
$$\frac{d_c}{d_1} = 1 - \frac{2l_1 p}{d_1 b} = 1 - \frac{K}{d_1 b} \dots \dots \dots (5, 1.3)$$

Where $K=2l_1 p$. For lamps with a given p and a particular positive diameter (d_1) the expression $2l_1 p/d_1$ is constant. The application of equation (5, 1.3) to positives 16 mm. in diameter and having various core sizes is shown in fig. 6. Points plotted here show the relation between d_c/d_1 and rate of burn b for individual readings, for the particular value of $K=960$ so that the curve passes through the point $d_c/d_1=0.8$, $b=300$. This point was taken as the mean of a number of readings at a current density about midway between the upper and lower limit at which the stable Beck effect occurs in the arc.

It will be observed that at burning rates in excess of about 200 mm. per hour, the experimental points occur in roughly equal numbers on either side of the calculated curve, but at lower burning rates there is a bias in the direction of a crater diameter larger than the calculated figure.

The experimental points in fig. 6 are derived from the data given by lamps A and B, and it appears that no significant difference can be detected between the two groups of results. Reference to Table I. shows that the value of p is not the same for the two lamps, but the constructional details of lamp B, which has the larger projection, were, in fact, such as to tend to delay the free access of oxygen to the surface of the burning carbon. This will result in a decrease in the value of l_1 so giving a fortuitous coincidence in the value K for the two lamps. Both of these lamps, however, were designed for the same practical application, for which the maximum crater diameter compatible with steady burning would be a desirable feature.

Fig. 6.



A corresponding plot of the data for lamp C gives results which are not inconsistent with those given in fig. 6 and with the conclusions drawn from them, but there is a much greater spread in the points, and so the agreement between experimental and calculated values is not as good as in the case of lamps A and B.

Several reasons are known why the results should be less consistent with large carbons than with small. In particular it is difficult to make the measurement d_c/d_1 accurately when a shallow crater in a large diameter carbon is concerned. Moreover, the differing thermal capacities of the carbons and the different times thereby involved in their reaching a stable condition after switching on, is one factor to which it was later realized that insufficient attention had been given.

As already implied, the value of d_c/d_1 at the low burning rates require special consideration, and in this connection there is one general conclusion which may be drawn from fig. 6: the agreement between experimental and calculated results at the higher burning rate suggests that, under these conditions the postulated mechanism is effective, particularly with the 16 mm. carbons, but the increased crater diameter found at low burning rates shows that oxidation under these conditions does not proceed so rapidly. As there is a monotonically increasing relation between burning rate and power input, it follows that the rate of oxidation is low for small power inputs and increases to an approximately constant figure for the larger power inputs. This, in its turn implies—what is known to be the case, as otherwise the carbon would continue to burn after the current was switched off—that the oxidation, though exothermic, is not self-supporting.

2. Evaporation from the Crater.

Dealing now with the effect of crater phenomena on the burning rate, it seems likely on general grounds that the rate of evaporation is measured by the crater area rather than by the carbon area. Fig. 6 lends support to this view, because it indicates that the decrease of internal diameter from d_1 to d_c is by a substantially self-supporting process such as oxidation.

If this is so, then the volume of material evaporated from the anode is

$$E = \frac{\pi}{4} d_c^2 b \text{ mm.}^3 \text{ per hour,} \quad \dots \dots (5, 2.1)$$

or, from equation (5, 1.3),

$$E = \frac{\pi}{4} b \left(d_1 - \frac{2l_1 p}{b} \right)^2 \text{ mm.}^3 \text{ per hour,} \quad \dots \dots (5, 2.2)$$

which gives E in terms of constants of the lamp and the burning rate. This would be a useful relation if there were any readings available of the actual rate of evaporation. Failing these, the best that can be done for verification is to assume that this rate of evaporation is closely related to the power input to the arc. If we assume, for example, that

$$E = mP + n, \quad \dots \dots (5, 2.3)$$

where m and n are constants, we find

$$P = \frac{\pi b}{4m} \left(d_1 - \frac{2l_1 p}{b} \right)^2 - \frac{n}{m},$$

$$P = \frac{\pi}{4m} \left(d_1^2 b + \frac{4l_1^2 p^2}{b} \right) - \frac{1}{m} (n + \pi l_1 p d_1), \quad \dots \dots (5, 2.4)$$

which is the relation shown in figs. 2, 3, and 4.

By comparing equation (5, 2.4) with curves drawn through the five sets of points in figs. 2, 3, and 4 and indicated in Table II., five sets of values of m and n are obtained. These also are given in Table II., together with the mean values.

There is a clearly qualitative agreement between the values of m and n deduced from the various sets of data, and with the exception of the

24 mm. carbon in lamp C, perhaps something better. It is, at any rate, not to be expected that complete accord should be obtained in a relation involving not only somewhat difficult measurements, but also the approximation of equation (5, 2.3).

With these reservations it is perhaps worth while to accept the mean values of m and n given by Table II. in order to examine the calculated curves corresponding, say, to those experimental values shown in fig. 4. We thus obtain, with these mean values m and n

$$P=0.0484(d_1^2b+4l_1^2p^2/b)+(1.14\times10^4-0.194l_1pd_1). \quad (5, 2.5)$$

TABLE II.

Lamp	d_1	m	n
A	16	12.5	-110×10^3
B	16	17.8	-161×10^3
C	16	18.5	-170×10^3
	20	12.2	-166×10^3
	24	20.0	-311×10^3
	Means	16.2	-184×10^3

The rate of burn, as a function of the power input to the lamp for the particular conditions (l_1 and p) appertaining to lamp C, calculated from the above equation is shown in fig. 4 and indicates a satisfactory explanation for the trend of the curves.

§ 6. CRATER DIAMETER AND DEPTH.

For the purpose of the present discussion, the shape of the crater was assumed to be a truncated cone with the measured crater diameter as the base and the core diameter as the smaller section. In fact, of course, the end of the core in the crater is slightly concave and the sides of the crater are also somewhat curved, but we are not able here to take these refinements into account.

There is perhaps little doubt that the explanation for the formation of the crater is to be found in the different rates of evaporation of the core and shell inside the crater itself.

We have already assumed a rate of recession, l_1 mm. per hour of the outer surface of the shell, due to oxidation. We now assume for the sloping sides of the inside of the crater a normal rate of recession of l_2 mm. per hour, mainly due to evaporation. The rate of recession of the surface of the core will, of course, be equal to the rate of burn, *i. e.* b mm. per hour. We therefore have (fig. 5)

$$\sin \theta = l_2/b, \quad (6.1)$$

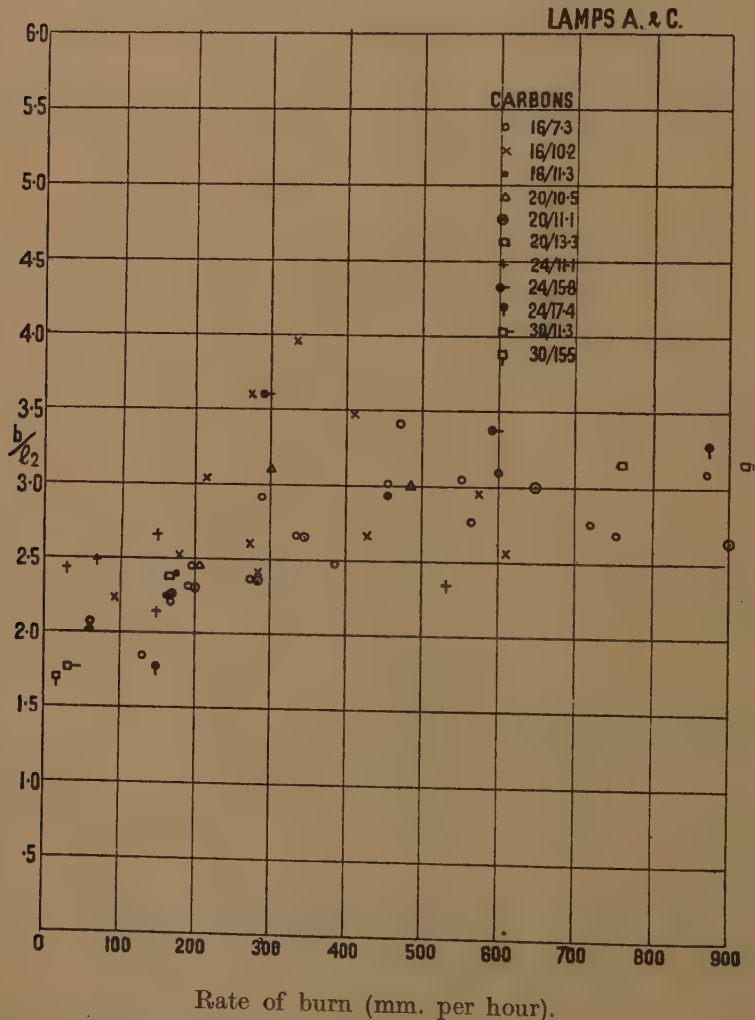
where θ is the angle of slope of the side of the crater. Now crater diameter is given by

$$d_c = d_1 - \frac{2l_1p}{b}$$

which follows from (5, 1.2) when

$$\cos \phi \doteq 1.$$

Fig. 7.



Therefore, crater depth is given by

$$\begin{aligned} h_c &= \left(\frac{d_1 - d_2}{2} - \frac{l_1p}{b} \right) \cot \theta, \\ &= \frac{1}{2}(d_c - d_2) \sqrt{\left(\frac{b}{l_2} \right)^2 - 1}. \quad \dots \dots (6.2) \end{aligned}$$

From this relation, experimental values of h_c , d_c and d_2 from Table II. allowed values of b/l_2 to be calculated. If the assumption that both b and l_2 are mainly the result of evaporation is correct, it would be expected that their ratio would be substantially constant, independent of the burning condition of the arc; provided, as in all the experiments considered here, the cores did not differ among themselves in composition, and the shells, although different from the cores, were uniform among themselves in chemical composition and physical state. The parameter chosen against which to plot b/l_2 was b , as this is probably the best single variable to indicate burning conditions. The results are shown in fig. 7. No data are included from lamp B, which was without positive rotation, since the irregularity of the crater edge formed in such circumstances, considerably reduces the accuracy of the determination of crater depth and may, in extreme circumstances, make the measurement meaningless. For very shallow craters and for burning conditions where $(d_c - d_2)$ becomes small, even approximate measurements are not possible, and such data have, therefore, also been omitted from fig. 7.

It will be seen that, over practically the whole of the range of b , the ratio b/l_2 is constant within the limits of experimental error, giving a value of θ of approximately 20° . At the smaller values of b , the angle θ appears to increase, although less rapidly for the smaller diameter carbons than for the larger. There seems to be two reasonable explanations for this, either or both of which may be operative. The division of current between core and shell may change so that a relatively larger proportion of the current flows to the crater through the shell, or the oxidation process which has previously been assumed only to apply to the external surface of the carbon, may also occur on the inner surface, for the distinction between external and internal surfaces largely disappears when the depth of crater is zero.

§ 7. ARC VOLTAGE.

The only measured voltage in experiments described here was the total voltage across the terminals of the lamp. This consists of the voltage drop along the positive and negative carbons and across the lamp mechanism, the cathode fall, the voltage drop in the negative flame and through the crater. We make the usual assumption that in normal operation the voltage drop in the arc is the major part of the total voltage across the lamp.

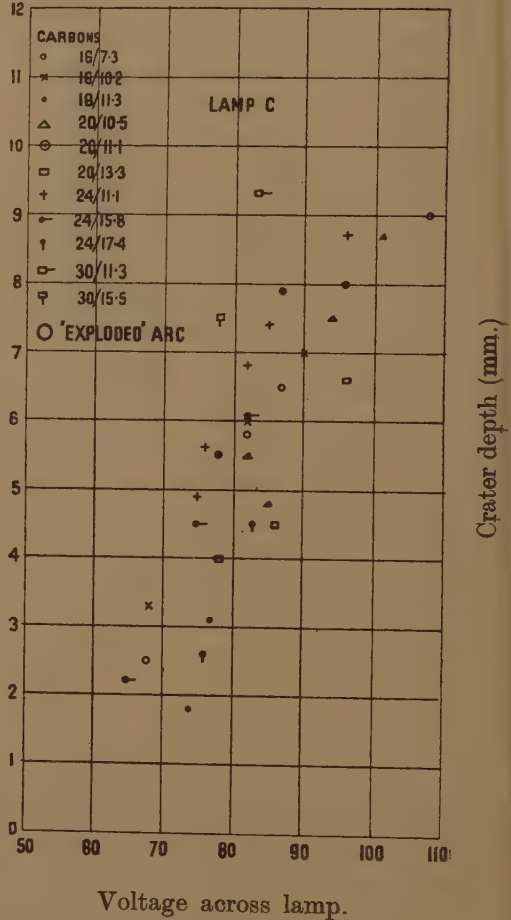
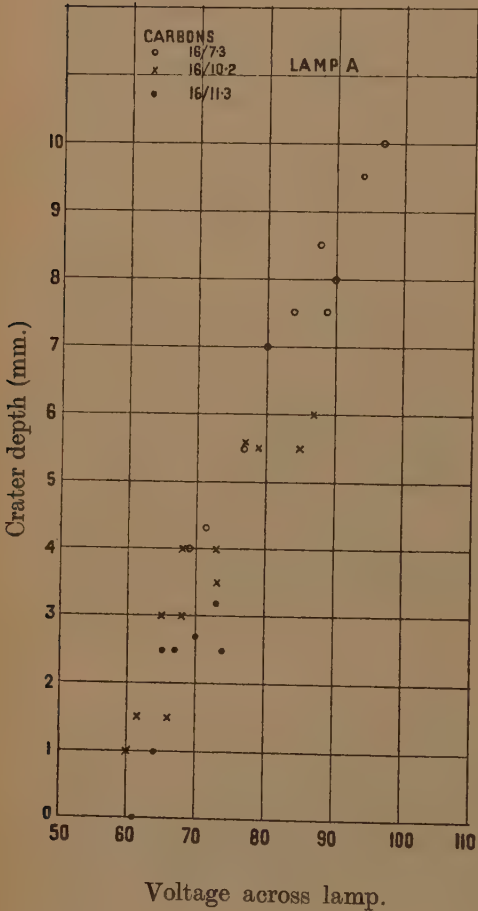
It is possible that the crater voltage is as low as 50 per cent of the total in which event practically the entire remainder would be accounted for by the cathode fall and the drop in the negative flame. So many factors are involved in the division of the voltage between the various parts of the arc that an attempt to disentangle them theoretically would be very formidable while the difficulties of an experimental investigation are well known. Nevertheless, it has been thought worth while to see if any information of value can be extracted from the measurements recorded here.

If we make the assumption, for which there is some theoretical justification, that the voltage drop in the negative flame, so long as the arc length is held constant, varies little for widely different conditions of burning; and that the cathode fall also varies little, we might consider an approximation of the form

$$V = \alpha h_c + \beta, \quad \dots \dots \dots (7.1)$$

Fig. 8 a.

Fig. 8 b.



where α and β are constants, and it might be expected that α and β are independent of d_1 or d_2 as well as of the burning conditions. Of course, β would be a measure of the voltage drop external to the crater and α a measure of the voltage drop per cm. of crater depth.

In figs. 8 a and 8 b are shown the results for lamps A and C respectively, and it is clear that an approximate relation of the form considered does, in fact, hold, where the constants have the same value for both lamps. It appears also that similar conclusions hold if the mean crater depth is used in place of the crater depth (h_c) with, of course, a different value of α .

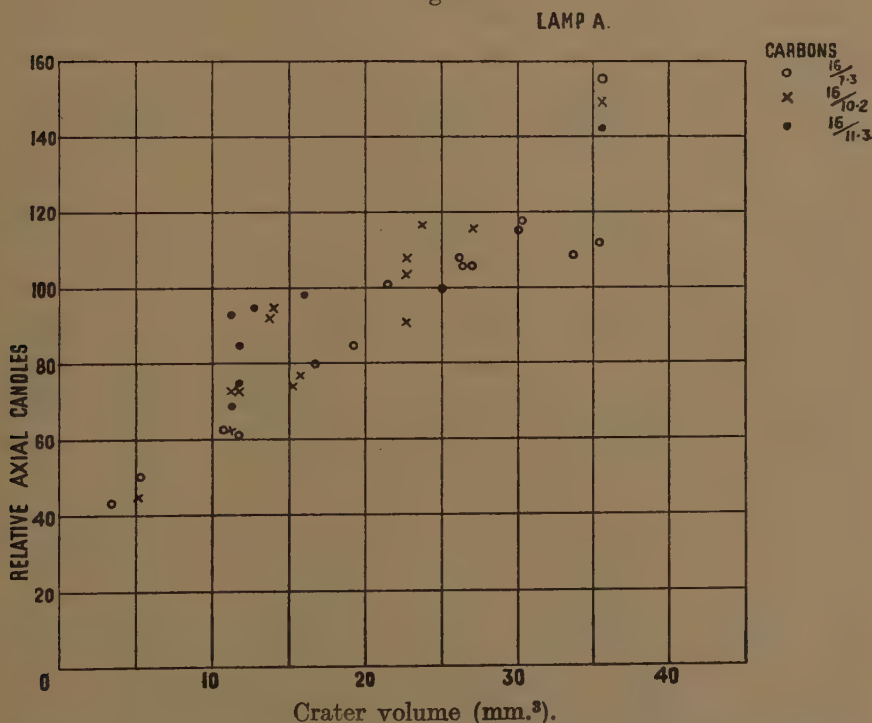
For this purpose the mean crater depth is taken as the volume of the truncated cone divided by the mean area of the inner and outer sections.

The figures do, in fact, show a remarkable approximation to the relation which can hardly be fortuitous. Taken on its face value it would indicate that there is a total voltage drop external to the crater of about 60 volts, and that the voltage drop through the crater depth (h_c) is of the order of 40 volts per cm. Moreover, these values appear to be constant over the wide range of carbons and operating conditions used in these experiments.

§ 8. LIGHT OUTPUT MEASUREMENTS ; RELATIVE AXIAL CANDLES.

If one can assume that the gases and vapours in the crater are at a substantially uniform temperature throughout its volume, and that the radiating gas and vapour in the crater are substantially transparent to

Fig. 9.



their own radiation, then the total light radiated through the crater will be proportional to its volume, as will also the axial candles.

Fig. 9 shows the relative axial candles for lamp A with the usual range of carbons, but, for reasons similar to those given in § 6, the results corresponding to crater depths less than 1.5 mm. were omitted. Within limits of experimental error it appears that the assumptions are justified. A similar plot for the results given by lamp C leads to the same conclusion, although the spread of the points is greater, with the exception that points corresponding to the exploded condition definitely do not fall on the same curve.

§ 9. CONCLUSIONS.

In the above sections the attempts to explain observed relations between various parameters lead naturally to the suggested equations between them. These equations should be valid for any conditions within the range considered in this paper, and one might expect that they would represent a much wider range of operation. If this is so, it is legitimate to consider these equations and the assumptions on which they are based as a tentative theory of the mechanism of formation and functioning of the positive crater. In these equations appear, of course, constants which must be considered as characteristic of the particular lamp and carbons used, for example, the power input to the lamp, P .

The positive projection, p , is a constant of a similar nature, specifying the lamp. The constants d_1 and d_2 specify the dimensions of the positive. The physical properties of the carbon are specified by l_1 and l_2 . The empirical constants m, n, α, β , have less determinable physical significance. The arc length, assumed constant, does not appear explicitly in the equations although the constants n, β , are related to it. From these data, it is possible to deduce all the essential properties of the positive crater and the rate of burn of the positive carbon, as well as the voltage across the lamp.

The process by which this is possible is fairly evident. Thus, for instance, the rate of burn is obtainable from equation (5, 2.5) or fig. 4; the crater diameter, d_c , then follows from equation (5, 1.3) or the corresponding curves of, for example, fig. 6; the crater depth, h_c , is given at once by equation (6.2), being proportional to $(d_c - d_2)$ over the range for which b/l_2 is constant (see fig. 7). For other values of b/l_2 the empirical values in fig. 7 must be used. From the crater dimensions thus determined, the mean crater depth can be calculated, and therefore the voltage drop across the arc from equation (7.1).

It is not suggested that this procedure provides a practical method of predicting with any accuracy the actual behaviour of an arc lamp. It does, however, provide a qualitative explanation of the arc phenomena, and, in fact, within limits, is capable of giving quantitative results. The range of operating conditions for which this is true is determined by the range over which equations (5, 2.3) and (7.1) hold and the range of validity of the assumptions made with regard to evaporation and oxidation.

The equations and assumptions on which the proposed mechanism of the arc depends should be capable of further refinement; for example, equations (5, 2.3) and (7.1) are probably approximations holding only over a limited region, and further experimental work would be needed to produce the more accurate relations. It has not been possible for us to do this work, but we thought that the observations and the discussion of them were of sufficient interest to be put on record in their present form.

REFERENCES.

- FINKELNBURG, W., 1940, *Zeit. f. Phys.*, **116**, 214.
Symposium on Searchlight. The Illuminating Eng. Soc., 1948.

XLIX. *Transit-time Deterioration of Space-charge Reduction of Shot Effect.*

By D. K. C. MACDONALD,
Clarendon Laboratory, Oxford*.

[Received November 30, 1948.]

§ 1. INTRODUCTION.

IT is well known that the emission current from either a thermionic or photo-electric cathode exhibits complete randomness ("full shot-effect") such that :

$$\overline{\delta i_f^2} = 2eI df, \quad (1)$$

where $\overline{\delta i_f^2}$ is the mean square fluctuation of the current observed within an arbitrary frequency interval df ; e is the electronic charge; I is the mean current.

When, however, for example, space-charge is present, the current drawn exhibits less fluctuation than expressed by equation (1) :

$$\overline{\delta i_f^2} = 2eI\Gamma^2 df \quad (0 < \Gamma^2 \leq 1), \quad (2)$$

Due to the interaction of the electrons through the medium of the space-charge, a certain degree of order has been introduced. If now this current drifts for some distance, as in a velocity-modulation valve, we should anticipate that a progressive disordering will re-establish itself until after a sufficient time the current will once more exhibit full shot noise ($\Gamma^2=1$). We wish then to solve the problem of the variation of Γ^2 as a function of this drift time, τ .

§ 2. APPROXIMATE SOLUTION.

Before analysing the general problem, let us derive an approximate solution for a case of practical significance. We consider a current that has been "smoothed" ($\Gamma^2 \rightarrow 0$) due to space-charge by the normal mechanism as in a diode valve; in this case if a negative fluctuation of current occurs, then the potential minimum is raised a little and a "compensating" current is allowed to pass, so minimizing the net fluctuation observed. The primary fluctuation may arise in electrons of any energy-class, but the compensating current will always be provided from electrons of practically zero energy, since it is just these electrons that are directly affected by small variations of the depth of the potential minimum. Let us assume that the total current after leaving the cathode (and potential minimum) is accelerated with negligible transit time to a relatively high velocity, V .

* Communicated by the Author.

and is then allowed to "drift" (as in practice) for a time τ until it reaches some frequency-sensitive device—say a rhumbatron. We then require to know what value of τ is permissible before serious increase of noisiness will occur in the electron beam.

The "primary" noise current, on leaving the potential minimum, will have an effective root mean square velocity u , the "compensating" component having practical zero velocity. After the acceleration, the velocities are given by

$$v_p \doteq V + \frac{1}{2} \frac{u^2}{V}; \quad v_c \doteq V \quad (u \ll V). \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Hence the difference, $\Delta\tau$, in time of arrival at the receiving plane, distant S , is given by

$$\begin{aligned} \Delta\tau &= -S \left\{ \frac{1}{v_p} - \frac{1}{v_c} \right\} \doteq S \frac{1}{2} \frac{u^2}{V^3} \\ &= \frac{1}{2} \tau \frac{u^2}{V^2}, \end{aligned}$$

i. e.

$$\tau = \frac{2\Delta\tau\Phi}{\epsilon}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where ϵ is the average energy of the primary component and Φ is the accelerating energy.

We might then suggest that if serious deterioration is to be avoided, the two currents must arrive within say 1/20 cycle of one another and hence

$$\tau = 0.1 \Phi / \epsilon \text{ cycles.} \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

If the electrons are emitted from a cathode of temperature $\theta \sim 10^3$ °K. then $\epsilon \doteq k\theta (\sim 0.1 \text{ eV})$ and if the accelerating voltage is about 100 volts, then $\tau \doteq 100$ cycles.

§ 3. GENERAL SOLUTION.

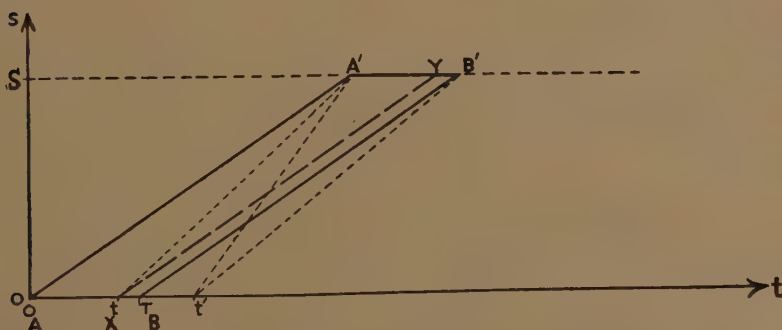
We now wish to consider the more general case of a beam of electrons derived from a cathode, temperature θ , which is entirely smoothed in numerical fluctuation ($I^2=0$) but which exhibits a random Maxwellian distribution, characteristic of the emitter temperature, and to derive the complete variation of I^2 with τ . In the language of a paper by R. Kompfner and the writer (1948) we should say that the beam has zero "shot noise" but full "chromatic noise". In the case of a physical beam smoothed through the action of the space-charge potential minimum this cannot be precisely true, since with this mechanism it is low-energy electrons that "fill the gaps" in the high energy current streams and thus a degree of correlation must, of course, exist in the energy-spectrum of the smoothed current. However, it is not difficult to see that the ideal case should prove a good approach to the physically realizable beam;

in addition, we can compare our solution with the foregoing approximate solution. The phenomenon is of the type akin to the "transient" Brownian-movement diffusion problem where we analyse the progression from the initial specified velocity, say, to the limiting "equilibrium" random distribution (*e. g.* Uhlenbeck and Ornstein (1930)).

3.1. Analysis.

We consider a series of electrons initially spaced regularly in time at a rate N per second. Let us represent the subsequent motion of the electrons on a distance-time diagram (fig. 1) as used by Applegate. An electron starting its drift at time t with just the "applied" velocity V will have its path delineated by the sloping line XY ; if its velocity be greater its path-line will be steeper. We wish to examine the variation with drift time of the number of electrons to be found in an interval of arbitrary length T , long enough to contain a relatively large number of

Fig. 1.



electrons. At 10^{10} c./s., for example, a current of 10^{-6} amps. provides $\sim 10^3$ electrons in a single cycle.

We have assumed that the electrons have a random energy component, say, E , in the direction of drift characteristic of the emitter temperature θ ;

i. e.

$$\left. \begin{aligned} p(E) dE &= (1/k\theta) \exp(-E/k\theta) dE, & (E > 0) \\ &= 0, & (E < 0) \end{aligned} \right\} \quad \dots (6)$$

Let τ be the drift time to the plane $s=S$ for an electron with $E=0$ (*i. e.* with velocity V). Then the velocity and transit time of an electron starting with a thermal component E are $V \cdot (1+E/2\Phi)$ and $\tau \cdot (1-E/2\Phi)$ (where $\Phi = \frac{1}{2}mV^2$) to the first order. We divide the electrons into two classes; those originally within the interval T (AB in fig. 1) and those outside that interval.

Then an electron of the former class whose incidence lies within $(t; t+dt)$ will *not* be found within the corresponding interval ($A'B'$) after drifting

for τ if $\tau E/2\Phi > t$, i. e. if $E > 2\Phi t/\tau$. Thus the probable loss of electrons from the group in drift time τ will be

$$\begin{aligned}\bar{n}_1 &= \int_0^T \left\{ \int_{2\Phi t/\tau}^{\infty} (1/k\theta) \exp(-E/k\theta) dE \right\} N dt, \\ &= (N\tau k\theta/2\Phi) \{1 - \exp(-2T\Phi/\tau k\theta)\}. \quad . \quad . \quad . \quad (7)\end{aligned}$$

Similarly the probable gain of electrons to the group from those originally outside the interval AB is given by

$$\begin{aligned}\bar{n}_2 &= \int_T^{\infty} \left\{ \int_{(t-T)2\Phi/\tau}^{2\Phi t/\tau} (1/k\theta) \exp(-E/k\theta) dE \right\} N dt^* \\ &= (N\tau k\theta/2\Phi) \{1 - \exp(-2T\Phi/\tau k\theta)\}. \quad . \quad . \quad . \quad . \quad (8)\end{aligned}$$

Since \bar{n}_2 and \bar{n}_1 are equal in magnitude it follows that the net average change is nil, as we should anticipate.

To compute the mean square fluctuation we bear in mind that if the probability of an event is p , then the probable mean square fluctuation is given by $\overline{\delta n^2} = npq$ ($q = 1 - p$) in the limit of n large (Bernoulli distribution). We thus obtain

$$\begin{aligned}\overline{\delta n_1^2} &= \int_0^T \{p(1-p)\} N dt, \quad \text{where} \quad p = \int_{2\Phi t/\tau}^{\infty} (1/k\theta) \exp(-E/k\theta) dE \\ &= (N\tau k\theta/4\Phi) \{1 - 2 \exp(-2\Phi T/\tau k\theta) + \exp(-4\Phi T/\tau k\theta)\}, \quad . \quad . \quad (9)\end{aligned}$$

and similarly,

$$\overline{\delta n_2^2} = (N\tau k\theta/4\Phi) \{1 - \exp(-4\Phi T/\tau k\theta)\}, \quad . \quad . \quad . \quad . \quad (10)$$

and therefore

$$\overline{\delta n^2} \equiv \overline{\delta n_1^2} + \overline{\delta n_2^2} = (N\tau k\theta/2\Phi) \{1 - \exp(-2\Phi T/\tau k\theta)\}. \quad . \quad . \quad (11)$$

This expression then determines the mean square fluctuation of numbers of electrons for arbitrary T and τ . We note immediately that for "small" τ ($\tau \ll T\Phi/k\theta$), $\overline{\delta n^2}$ is proportional to τ and independent of T . This is in agreement with intuition, which suggests that under this condition only electrons very close to the ends of the interval T can contribute significantly to the fluctuation. On the other hand, as $\tau \rightarrow \infty$ we have

$$\overline{\delta n^2} = NT, \quad . \quad . \quad . \quad . \quad . \quad (12)$$

the familiar statistical equation indicative of complete randomness. If we denote the current fluctuation averaged over the interval T by δi_T , then

$$\overline{\delta i_T^2} = \frac{e^2 \overline{\delta n^2}}{T^2} = \frac{eI\tau k\theta}{2\Phi T^2} \{1 - \exp(-2T\Phi/\tau k\theta)\}. \quad . \quad . \quad (13)$$

* The existence of infinity as upper limit of the outer integral here is justified by the rapid convergence of the expression $p(E) dE$, and is equivalent to the assumption: $E \ll \Phi$, used in deriving the expressions for velocity and transit time.

This is to be regarded as a statistical equation from which we wish finally to deduce the corresponding (power) frequency spectrum: $w(f)$. In the Appendix (equation (A.6)) it is shown that this may be determined by means of the following inversion theorem:—

$$w(f) = 4\pi f \int_0^\infty \frac{\partial \Psi}{\partial T} \sin 2\pi f T dT \quad . \quad . \quad . \quad (14)$$

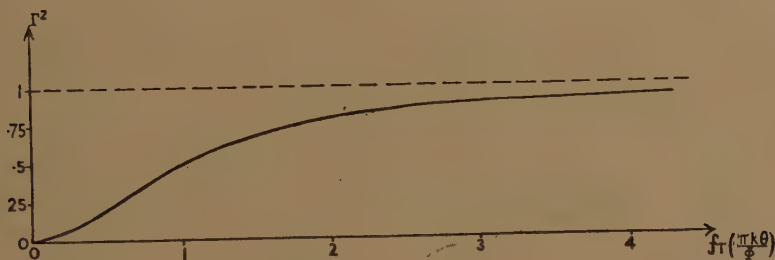
(where $\Psi(T) = T^2 \cdot (\overline{\delta i_T^2})$); thus

$$w(f) = 4\pi f e I \int_0^\infty \exp(-2\Phi T / \tau k \theta) \cdot \sin 2\pi f T dT,$$

$$\overline{\delta i_f^2} \equiv w(f) df = 2eI df \left\{ \frac{4\pi^2 f^2 \tau^2}{4\pi^2 f^2 \tau^2 + (2\Phi/k\theta)^2} \right\} \quad . \quad . \quad . \quad (15)$$

The factor in brackets may, of course, be identified with Γ^2 (cf. equation (2)), and it is immediately evident that for $\tau=0$, $\Gamma^2=0$; $\tau \rightarrow \infty$, $\Gamma^2=1$. Γ^2 is intrinsically a function of $f\tau$, i. e. the number of “drift cycles”, as we

Fig. 2.



might expect, and of the energy ratio $\Phi/k\theta$. A sketch of Γ^2 as a function of $f\tau$ is shown in fig. 2, and fig. 3 is intended to illustrate how the fluctuation frequency spectrum builds up as the drift time increases.

If we take $\theta \sim 10^3$ °K and an accelerating potential ~ 100 volts, then Γ^2 will rise from 0 to ~ 0.1 in ~ 100 cycles.

§ 4. LIMIT OF EXTREMELY HIGH FREQUENCIES AND VERY SMALL CURRENTS.

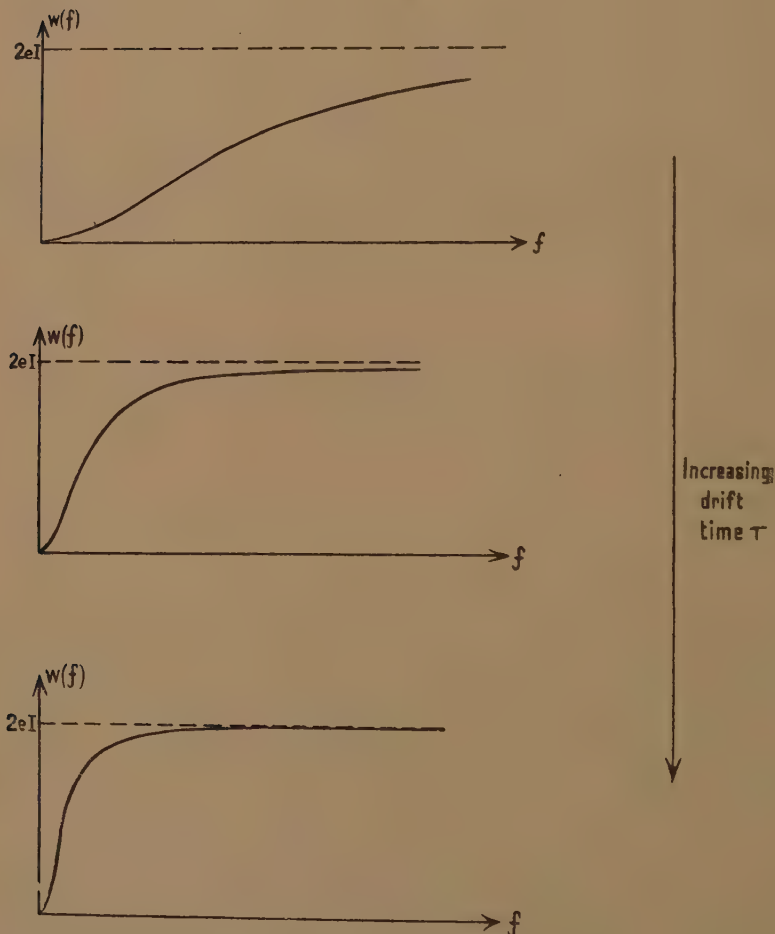
If the current is so small and/or the frequency so high that the significant time-interval, T , can no longer be considered to contain “many” electrons, then it is evident that the above analysis is no longer applicable; the situation seems unlikely to arise, at least in the near future, but an analysis has been made of this limiting case, based on a consideration of the fluctuation in the time-interval between any two successive electrons. We might perhaps say that we are examining a “fine-structure” effect under these circumstances.

In this case it appears that the spectrum of the fluctuation as it develops is always effectively uniform (“white”); therefore we can legitimately split up the current, I , at any stage of the process into a “noisy” component $I\Gamma^2$, and a “smooth” component $I(1-\Gamma^2)$, where the former may

be regarded as entirely random. A differential equation for Γ^2 can thus be set up—entirely analogous to the diffusion equation mentioned in 3.1 above—and the final solution of the problem rests in a determination of the initial “slope” $(d\Gamma^2/d\tau)_{\tau=0}$. It is found that

$$\left(\frac{d\Gamma^2}{d\tau}\right)_{\tau=0} = \frac{\sqrt{6I}}{e} \sqrt{\frac{3k\theta}{\epsilon\Phi}} \quad (\epsilon=2.718\dots), \quad \dots \quad (16)$$

Fig. 3.



and the differential equation for Γ^2 reads

$$\frac{d\Gamma^2}{d\tau} = AI(1 - \Gamma^2)^2,$$

whence

$$\Gamma^2 = \frac{AI\tau}{1 + AI\tau} \quad \dots \quad (17)$$

where $A = \sqrt{18k\theta/(\epsilon \cdot e\Phi)}$, and assuming $\Gamma^2=0$, $t=0$.

§ 5. ACKNOWLEDGMENTS.

The writer wishes to express his sincere thanks to Mr. R. Kompfner for much valuable and stimulating discussion. Thanks are due also to Mr. D. F. Gibbs for valuable comments on the analysis.

APPENDIX.

Let us adopt for analysis the statistical representation of "noise" employed by Rice (1944, 1945), based on similar expressions used earlier by Rayleigh, Schottky and others:

$$i(t) = \sum_n (a_n \cos \omega_n t + b_n \sin \omega_n t), \quad (A.1)$$

where a_n , b_n are normally and independently distributed such that $\overline{a_n^2} = w(f_n) \Delta f$ where Δf represents the elementary interval between f_n and f_{n+1} .

We can then readily show that

$$\overline{\delta i_T^2} \equiv \left\{ \frac{\left(\int_t^{t+T} i(t) dt \right)^2}{T^2} \right\}^t = 2 \int_0^\infty \frac{w(f)}{4\pi^2 f^2 T^2} (1 - \cos 2\pi f T) df. \quad . . (A.2)$$

It may be verified from equation (A.2) immediately that a "white" spectrum ($w(f) = 2eI$) leads to $\overline{\delta i_T^2} = eI/T$, the equation originally derived by Schottky (1918) for the "shot effect". Setting $\Psi(T) = (T^2 \cdot \overline{\delta i_T^2})$ and remembering that $w(f)$ is an even function of f , we have

$$\Psi(T) = \int_{-\infty}^\infty \frac{w(f)}{4\pi^2 f^2} (1 - \cos 2\pi f T) df, \quad (A.3)$$

$$\therefore \Psi'(T) = \int_{-\infty}^\infty \frac{w(f)}{2\pi f} \sin 2\pi f T df. \quad [\Psi'(T) \equiv \partial \Psi / \partial T], \quad . . (A.4)$$

assuming the usual requirements of convergence, etc., to be satisfied, *i. e.*

$$\Psi'(T) = \frac{1}{i} \int_{-\infty}^\infty \left(\frac{w(f)}{2\pi f} \right) \cdot \epsilon^{2\pi i f T} df.$$

Hence

$$\begin{aligned} \frac{w(f)}{2\pi f} &= i \int_{-\infty}^\infty \Psi'(T) \cdot \epsilon^{-2\pi i f T} dT \\ &= 2 \int_0^\infty \Psi'(T) \cdot \sin 2\pi f T dT, \quad (A.5) \end{aligned}$$

since $\Psi'(T)$ is an odd function of T (by (A.4)), and again assuming convergence, etc., *i. e.*

$$w(f) = 4\pi f \int_0^\infty \Psi'(T) \cdot \sin 2\pi f T dT, \quad (A.6)$$

the result quoted in the text.

It is felt that this method of approach should be of value in a number of stochastic problems since, for example, it avoids the complication of Dirac functions to represent individual electron events.

REFERENCES.

- MACDONALD, D. K. C., and KOMPFFNER, R., 1948, *Quantum effects in the interaction of electrons with high-frequency fields*
 RICE, S. O., 1944, *Bell. Syst. Tech. Jour.*, **23**, 282; 1945, *Ibid.*, **24**, 46.
 SCHOTTKY, W., 1918, *Ann. Phys., Lpz.*, **57**, 541.
 UHLENBECK, G. E., and ORNSTEIN, L. S., 1930, *Phys. Rev.*, **36**, 823.

L. *The Efficiency of the Selenium Barrier-Photocell when used as a Converter of Light into Electrical Energy.*

By E. BILLIG, Dr.tech., and K. W. PLESSNER, M.Sc., Ph.D.,
 Research Laboratory, Associated Electrical Industries, Aldermaston*.

[Received February 4, 1949.]

1. INTRODUCTION.

IN a recent investigation on the efficiency of Se photocells when converting light into electrical energy, R. A. Houston (1948) concluded that "the highest efficiency found in this paper for approximately monochromatic light is 6.4×10^{-8} ; the highest efficiencies recorded for white light by two different arrangements were 2×10^{-6} and 6.2×10^{-5} , respectively".

These values seemed to us very low indeed. Taking performance figures from available pamphlets on photocells, simple arithmetic led us to expect efficiencies of the order of about 1 per cent. Furthermore, it seemed altogether wrong that monochromatic light of proper wavelength should, in this respect, give poorer results than white light, considering that Se photocells have their maximum sensitivity somewhere in the visible range. We therefore decided to look more closely into this question.

2. EXPECTED ORDER OF MAGNITUDE OF EFFICIENCY

A beam of monochromatic light—of wavelength λ or frequency $\nu = \lambda/c$ and of intensity J —is equivalent to a stream of photons, each of energy $E = h\nu = hc/\lambda$, at a rate $n (= J/h\nu)$ per unit cross-sectional area and per unit time. By a mechanism, the main features of which are by now fairly well understood (Lehovec 1948), but which need not be further discussed here, this stream of photons produces, on arrival at the cell, a number of current carriers in it and raises them to a potential (V) sufficient to force a current (I) through the load resistance (R) connected across its

* Communicated by the Authors.

terminals. The power output ($P=V \times I$) would obviously attain an absolute maximum if

- (i) the rate of electrons passing through the load were equal to the rate of photons arriving at the cell, *i. e.* if the "quantum yield" * $\eta_i=1$, and if, simultaneously,
- (ii) the potential difference (V) across the terminals of the cell were equal to the equivalent voltage (v , see above), *i. e.* if the "voltage yield" $\eta_v=V/v=1$.

In these circumstances, the efficiency η of converting an incident radiation into electrical energy would be equal to 1. For reasons set out by Lehovec (1948), neither of these conditions is fulfilled in practice and the conversion efficiency $\eta=\eta_i\eta_v$ never reaches unity.

Approximate values of both yields, η_i and η_v , can readily be taken from manufacturers' pamphlets giving guaranteed figures of current- and voltage-output. With a relatively low resistance in circuit, the photocurrent is directly proportional to the illumination. A good average value for the current under such conditions is $500 \mu\text{A/lumen}$. The open-circuit voltage (V_0), on the other hand, varies only slowly with the intensity of the illumination. It may reach 0.3 volt, but in poor light may drop to about 0.1 volt.

For green light ($\lambda=5500 \text{ \AA.U.}$, $v=2.23 \text{ eV.}$) the mechanical equivalent of light is 1.51 mW/lumen , and the figure of $500 \mu\text{A/lumen}$ corresponds to a quantum yield of

$$\eta_i = \frac{500 \times 10^{-6} \times 2.23}{1.51 \times 10^{-3}} = 0.75,$$

not far short of the theoretical limit. The voltage yield (η_v) is rather poorer, about 0.05 to 0.15, according to the intensity of the incident radiation (corresponding to the value of V_0 given above).

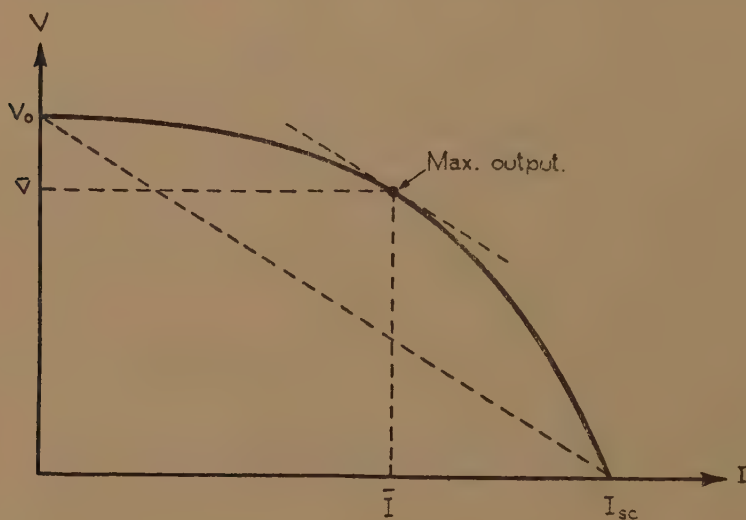
However, under either of these conditions, *i. e.* short-circuit or open-circuit, no power is taken from the cell. On load, *i. e.* with an external resistance connected across its terminals, the photocell exhibits the well-known non-linear characteristic illustrated in the figure. For a given illumination, maximum output is obtained when the load resistance is approximately equal to the ratio of open-circuit voltage (V_0) to short-circuit current (I_{sc}), which ratio may be termed "internal resistance" r of the cell. In this case both the current (\bar{I}) through, and the voltage (\bar{V}) across, the load are rather less than under short-circuit or open-circuit conditions, respectively. The actual values depend somewhat on the curvature of the characteristic, but the ratios \bar{I}/I_{sc} and \bar{V}/V_0 are usually of the order of 0.6 each; if the characteristics were linear, both these values would be 0.5, as is well known for a circuit with constant internal resistance.

* *i. e.* "quantum yield" is referred here to the number of incident and not of absorbed photons.

The efficiency of power conversion may thus be written

$$\eta = \frac{P}{n.e.v} = \eta_i \cdot \eta_v \cdot \bar{I}/I_{sc} \cdot \bar{V}/V_0,$$

and, substituting the values above, may be expected to reach about 1 to 4 per cent in the case of monochromatic illumination, at a wavelength near the peak sensitivity of the cell; somewhat less for white light, and especially so for light from an incandescent filament containing rather a lot of infra-red radiation to which the cell is not sensitive.



V/I characteristic of barrier photo-cell
under constant illumination.

3. MEASUREMENTS.

The efficiency of a commercial photocell of sensitive area 1.5 sq.cm. was measured by comparison with a Hilger-Schwartz thermopile. The maximum power output of the photocell under a given illumination was determined by varying an external resistance R until I^2R was a maximum. The calibration of the thermopile in terms of total incident radiation was known and its neutrality over the spectrum appears to be well established, at least up to wavelengths of several microns.

(a) *Monochromatic Light.*

A Hilger-Barfit wavelength spectrometer was used to isolate the most intense lines in the light from a 250 W. Hg arc lamp, and Table I. shows the results for the blue, green, and yellow lines. The intensities obtained after passage through the spectrometer were not very high, and in order to have an intensity comparable with that from an incandescent lamp or with daylight the mercury green line was isolated

by means of an Ilford filter, No. 807. A liquid filter ensured that no infra-red radiation was present. The result for this measurement is included in Table I.

TABLE I.

Wavelength (A.U.)	Total incident energy (μ W/sq.cm.)	Short- circuit current (μ A)	Open- circuit voltage (mV)	Max. power output (μ W)	Quantum yield (%)	Power efficiency (%)
4358	52.5	12.1	130	0.61	44	0.78
5461	144	36.2	200	2.65	38	1.22
5770, 5790	177	43.5	200	3.18	35	1.20
5461 (filter)	730	164	307	15.4	34	1.41

It is seen that the quantum yield is rather low, only about half of what might be expected for a good commercial cell. The values of the power efficiency could therefore be almost doubled if a better cell had been chosen.

(b) *Daylight and Incandescent Light.*

On a dull January morning a direct comparison between thermopile and photocell was made, the acceptance angle being limited to about 0.41 steradian for the purely practical reason of excluding draughts

TABLE II.

Radiation	Total incident energy (μ W/sq.cm.)	Short- circuit current (μ A)	Open- circuit voltage (mV)	Max. power output (μ W)	Power efficiency (%)
Daylight (overcast sky)	1,350	166	310	19.2	0.95
Direct sunlight (26th Jan., noon)	54,700	935	526	150	0.183
100 W. lamp	5,380	55.5	220	4.78	0.059

from the thermopile. The result is shown in Table II. The value of 0.95 per cent for the efficiency seems plausible since the white light scattered by the clouds consists almost exclusively of visible radiation to which the photocell, like the eye, is sensitive.

In contrast with this, the radiation from an incandescent lamp consists largely of infra-red radiation, to which the photocell is not sensitive. The results obtained with a 100 watt gas-filled bulb at a distance of 30 cm. are also given in Table II. The light intensity may be estimated from the

wattage of the lamp as 0.12 lumens/sq.cm., giving an output of $300 \mu\text{A/lumen}$. The fact that the power efficiency is 20 times less than with the other sources is, of course, related to the poor efficiency of an incandescent lamp as a source of light energy.

Similar results were obtained from direct sunlight on a clear day in January. With the sun low in the sky, visible radiation is absorbed much more strongly by the atmosphere than the infra-red, with a corresponding decrease in the power efficiency. The value shown in Table II. is intermediate between the light from the overcast sky and from the incandescent lamp. The total radiation recorded by the thermopile corresponds to about half of the sun's radiation arriving at the upper atmosphere.

4. CONCLUSIONS.

It will be seen that the measured efficiencies agree with the estimated value, which is not surprising since the latter is based on experimental results. The lowest value, obtained with an incandescent source, is still ten times greater than Houston's highest, and it must be concluded that—

- (1) Houston used decidedly inferior cells.
- (2) The load resistance in Houston's experiments was far below the optimum required for maximum output (*i. e.* of the order of the internal resistance of the cell).
- (3) To account for the very poor performance he observed with "monochromatic" light, we suspect that the filter used by him was not opaque to infra-red radiation, and what was measured was not the response to monochromatic visible light but mainly to infra-red.

Thanks are due to Dr. T. E. Allibone, F.R.S., for permission to publish this paper.

REFERENCES.

- HOUSTON, R. A., 1948, *Phil. Mag.*, **39**, 902.
LEHOVEC, K., 1948, *Phys. Rev.*, **74**, 463.

LI. *The Zero-Point Energy of a System of Particles.*

By R. B. DINGLE,

Royal Society Mond Laboratory, Cambridge *†

[Received March 15, 1949.]

ABSTRACT.

A variational method is used to estimate the zero-point energy of an assembly of particles satisfying Einstein-Bose statistics.

1. INTRODUCTION.

SEVERAL attempts have been made in the past to estimate theoretically the zero-point energy of a gas of hard spherical particles (diameter a), given the mean inter-particle distance l . In this paper a further calculation is made of this quantity.

The results of this investigation may be applied to the case of Helium II (Mott 1949), since some theories of the peculiar behaviour of this liquid assume the existence of a Bose-Einstein condensation—*i. e.* a “condensation” of a large fraction of the available atoms into the state of least energy (London 1938, 1939, Fowler and Jones 1938, Dingle 1949).

Two limiting cases have been examined in the literature, that in which $l \sim a$, when the mean interparticle distance is comparable with the particle diameter (London 1936), and secondly the case $l \gg a$, when the mean inter-particle distance is much larger than the particle diameter (Heitler 1927, Lenz 1929).

In the first of these cases a simple physical interpretation of the zero-point energy may be given; for if the density is so high that the particle diameter becomes comparable in magnitude with the inter-particle distance, each particle will become trapped within a box formed by its neighbours, and hence will acquire the zero-point energy appropriate to a confined particle. It is thus clear that the zero-point energy will increase with the density, since each particle will be confined within a smaller region. A rough estimate of the energy per particle may be made if we assume that each particle (of mass m) is confined within a spherical box of radius $l - a$. Then the Schrödinger equation for the problem may be written,

* The bulk of this work was carried out at the H. H. Wills Physical Laboratory, Bristol.

† Communicated by the Author.

using polar coordinates and remembering that the wave-function will be spherically symmetrical,

$$\frac{d^2}{dr^2}(r\psi) + \frac{8\pi^2m}{h^2}E(r\psi) = 0, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where E is the zero-point energy of the particle with wave-function ψ . The solution satisfying the boundary condition $\psi=0$ when $r=l-a$ is $r\psi=\sin \pi r/(l-a)$, with $E=h^2/8m(l-a)^2$.

It is much more difficult to give a correspondingly simple physical argument for the limiting case in which the particle diameter is much smaller than the mean inter-particle distance—*i. e.* for which $l \gg a$. Attempts have been made to estimate the zero-point energy in this case by Heitler (1927) and by Lenz (1929). Both methods involve the concept of self-consistent fields, but it is by no means obvious that such an approximation adequately replaces what is in reality a very complicated problem involving a multitude of boundary conditions, especially as clear physical interpretations of the arguments are lacking. In this paper we shall confirm the form of the expressions given by Heitler and Lenz and give an estimate of the numerical constants which arise.

2. THE CONSTRUCTION OF THE WAVE-FUNCTION.

We shall suppose that we are dealing with a system of hard spherical particles, each of diameter a , the density being such that the mean inter-particle distance is l . If the particles obeyed Fermi-Dirac statistics, the zero-point energy resulting from the Pauli Exclusion Principle would be of overwhelming importance, and it would be unnecessary to calculate the small corrections caused on account of the finite size of the particles. In the following, we shall therefore suppose that the particles obey Bose-Einstein statistics, and we shall study the problem by constructing a wave-function containing arbitrary parameters, and determine these by means of the variational principle.

The chosen wave-function must possess the following properties:—

- (a) It must be symmetrical in the coordinates of all the particles, since these obey Bose-Einstein statistics.
- (b) It must vanish whenever any two particles approach each other to within a distance a , since it is impossible that the distance between any two hard particles should be less than the particle diameter.
- (c) The wave-function must have no other zeros than those enumerated by condition (b), since we are seeking that wave-function which gives the lowest energy for the system. Wave-functions representing excited states would possess other zeros in addition.
- (d) When all the particles are widely separated, the wave-function should have a uniform value, unity say. This is a consequence of the assumption that the only forces of importance are those which come into play when two particles collide.

- (e) It should contain at least one adjustable parameter, so that the variational principle may be applied.
- (f) If the system of particles be confined within a box, the wave-function must vanish whenever any particle reaches the boundary. For at the boundary the potential would be infinite, so that the Schrödinger equation could not be satisfied unless the wave-function became zero.

In the following, we shall construct a one-parameter wave-function possessing these properties. We shall confine our attention to volume effects, so that the condition (f) may be ignored.

Let x_i, y_i, z_i be the Cartesian coordinates of a particle labelled i . For brevity, we write $x_{ij} = |x_i - x_j|$, etc., and $r_{ij}^2 = x_{ij}^2 + y_{ij}^2 + z_{ij}^2$, so that x_{ij}, y_{ij}, z_{ij} and r_{ij} are always positive. We take for the wave-function of the system of N particles obeying Bose-Einstein statistics, the product $\psi = \Pi f(r_{ij})$, this product being taken over the $\frac{1}{2}N(N-1)$ possible pairs of particles. In this expression, f represents some function which possesses the following properties:—

1. It vanishes when its argument r_{ij} becomes as small as a .
2. It possesses the asymptotic value unity for large values of the argument.

For instance, we may take

$$f(r_{ij}) = 1 - \exp[-\beta(r_{ij} - a)]; \quad r_{ij} \geq a; \quad \beta \geq 0.$$

This wave-function ψ then clearly obeys the conditions (a) to (e).

3. THE CALCULATION OF THE ENERGY.

Writing the Schrödinger equation

$$\nabla^2 \psi + \frac{8\pi^2 m}{h^2} E \psi = 0; \quad (\psi \text{ real}),$$

where $\nabla^2 = \sum_i \nabla_i^2$; $\nabla_i^2 = \partial^2 / \partial x_i^2 + \partial^2 / \partial y_i^2 + \partial^2 / \partial z_i^2$, we obtain

$$E = - \frac{h^2}{8\pi^2 m} \int \psi \nabla^2 \psi \, dV / \int \psi^2 \, dV,$$

where E is the energy of the whole system. Here the differential dV has been written for the product $\Pi dx_i dy_i dz_i$, and the integral is to be taken over all regions of space to which the particles have access. We easily see that $\partial f(r_{ij}) / \partial x_i = (x_i - x_j) f'_{ij} / r_{ij}$ where $f'_{ij} = \partial f(r_{ij}) / \partial r_{ij}$.

Of the $\frac{1}{2}N(N-1)$ factors in the product, $N-1$ will involve particle t , these being those factors formed from the combination of particle t with the $N-1$ other particles of the system. The operator will leave untouched the remaining factors, but for each of these factors which involve t , the wave-function will be multiplied, as a result of the operation, by $\{(x_t - x_q) / r_{tq}\} \{f'_{tq} / f_{tq}\}$ where $q \neq t$. Hence

$$\frac{\partial \psi}{\partial x_t} = \psi \sum_q \frac{x_t - x_q}{r_{tq}} \frac{f'_{tq}}{f_{tq}}.$$

There will be $N-1$ terms in this summation. Applying the operator $\partial/\partial x_i$ a second time, we obtain

$$\frac{\partial^2 \psi}{\partial x_i^2} = \psi \left[\sum_q \left\{ \left(\frac{x_{iq}}{r_{iq}} \right)^2 \frac{f''_{iq}}{f_{iq}} + \frac{f'_{iq}}{f_{iq}} \left(1 - \left(\frac{x_{iq}}{r_{iq}} \right)^2 \right) \right\} \right. \\ \left. + 2 \sum_{u,v} \frac{(x_i - x_u)(x_i - x_v) + \dots \cdot f'_{iu} f'_{iv}}{r_{iu} r_{iv} f_{iu} f_{iv}} \right],$$

the last summation being taken over all possible different pairs, u, v . The factor 2 arises since each pair u, v appears twice if both u and v separately range over their possible values. There will be $\frac{1}{2}(N-1)(N-2)$ different terms in this summation.

On adding the contributions due to the operators $\partial^2/\partial y_i^2$ and $\partial^2/\partial z_i^2$, and summing over t , we obtain

$$\nabla^2 \psi = \sum_t \nabla_t^2 \psi = 2\psi \left[\sum_{t,q} \frac{f''_{tq} + 2f'_{tq}/r_{tq}}{f_{tq}} + \sum_{t,u,v} \frac{(x_t - x_u)(x_t - x_v) + \dots \cdot f'_{tu} f'_{tv}}{r_{tu} r_{tv} f_{tu} f_{tv}} \right],$$

where both summations are taken over all the different possible terms. The factor 2 in the first summation arises since each pair t, q appears twice if both t and q separately range over their possible values. No further factor appears in the second summation, since no repetitions will be recorded on summing over t , because $u \neq v$ and $t \neq u, t \neq v$. There are $\frac{1}{2}N(N-1)$ different terms in the first summation, and $\frac{1}{2}N(N-1)(N-2)$ in the second.

Since $u \neq v$, and u and v are otherwise independent, the integral $\iint (x_t - x_u)(x_t - x_v) A(u) A(v) dx_u dx_v$ will vanish whenever $A(u)$ is an even function of x_u , or $A(v)$ of x_v . Since $r_{tu}, f(r_{tu})$ and $f'(r_{tu})$ are all functions of x_u^2 , we see that each of the terms of the second summation give zero contribution to the energy. Each term in the first summation will give a similar contribution, so that we may replace the summation by a typical term, *e. g.* that involving particles 1. and 2., multiplied by $\frac{1}{2}N(N-1)$, the number of terms in the summation. Hence the energy per particle, *e. g.* say, is given by

$$\epsilon = \frac{E}{N} = - \frac{(N-1)\hbar^2}{8\pi^2 m} \int \Pi_{i,j} f_{ij}^2 \left(\frac{f''_{12} + 2f'_{12}/r_{12}}{f_{12}} \right) dV \bigg/ \int \Pi_{i,j} f_{i,j} dV.$$

With the exception of the differentials, all the functions in this expression involve only the relative coordinates of the particles. To transform the differentials, we take as independent variables the set

$$x = \sum_i x_i / N; \quad y = \sum_i y_i / N; \quad z = \sum_i z_i / N.$$

which define the centre of gravity of the system, together with the coordinates of $N-1$ of the particles relative to the particle labelled 1; *i. e.* the set $x_{1i} = x_i - x_1$; $y_{1i} = y_i - y_1$; $z_{1i} = z_i - z_1$. The other relative coordinates, *e. g.* x_{ij} , are given in terms of these by expressions of the form $x_{ij} = x_{1i} - x_{1j}$. It may then be shown that the modulus of the

Jacobian of this transformation is unity. Also, since only the relative coordinates appear in the functions involved in E , the integrals over the coordinates of the centre of gravity are identical in numerator and denominator, and cancel. Hence we may replace dV by $\prod_i dx_i dy_i dz_i$.

By reducing the integrals to dimensionless forms and minimizing with respect to the parameters, it may be shown that ϵ is proportional to $(N-1)\hbar^2 a/m[V-a^3 \times \text{constant}]$, where a is the diameter of each particle. In the limiting case of a large number of very small particles

$$[V-a^3 \times \text{constant}] \rightarrow l^3$$

the cube of the mean inter-particle distance, and hence $\epsilon \propto \hbar^2 a/ml^3$.

4. ESTIMATION OF THE NUMERICAL CONSTANTS.

In this section we assume a specific type of wave-function containing one arbitrary parameter, and make a rough calculation of the zero-point energy per particle by means of the variational principle. We take $f(r) = 1 - \alpha \exp(-\beta r)$; $r \geq a$; $\alpha = \exp(\beta a)$ so that $\psi = \prod_{ij} [1 - \alpha \exp(-\beta r_{ij})]$, and $f'(r) = \alpha \beta \exp(-\beta r)$; $f''(r) = -\alpha \beta^2 \exp(-\beta r)$. Hence

$$f''_{12} + \frac{2f'_{12}}{r_{12}} = \alpha \beta^2 \exp(-\beta r_{12}) \left\{ 1 - \frac{2}{\beta r_{12}} \right\}.$$

Using the result of section 3, we have

$$\epsilon = \frac{(N-1)\hbar^2 \alpha \beta^2}{8\pi^2 m} \frac{\int \exp(-\beta r_{12}) (1 - 2/\beta r_{12}) [1 - \alpha \exp(-\beta r_{12})] \prod [1 - \alpha \exp(-\beta r_{ij})]^2 dV}{\int \prod [1 - \alpha \exp(-\beta r_{ij})]^2 dV},$$

where the products over i, j exclude i, j equal to 1, 2.

Taking into account only those contributions due to binary encounters (*i. e.* neglecting the cases in which another relative coordinate becomes small in addition to r_{12}), this reduces to

$$\epsilon = \frac{(N-1)\hbar^2 \alpha \beta^2}{8\pi^2 m} \frac{\int_a^\infty \exp(-\beta r) \cdot (1 - 2/\beta r) [1 - \alpha \exp(-\beta r)] 4\pi r^2 dr}{\int_a^\infty [1 - \alpha \exp(-\beta r)]^2 4\pi r^2 dr}.$$

It is easily shown that

$$\alpha \beta^2 \int_a^\infty [1 - \alpha \exp(-\beta r)] \exp(-\beta r) \cdot (1 - 2/\beta r) r^2 dr = 1/2\beta + \alpha^2 \beta + a,$$

on putting $\alpha = \exp(\beta a)$. Neglecting the small dependence of the denominator on β , we minimize $1/2\beta + \alpha^2 \beta + a$; the minimum occurs when $a^2 \beta^2 = \frac{1}{2}$, the expression then assuming the value $2.41a$. Thus for a large number of very small particles, the energy per particle is

$$\epsilon = \lim_{N \rightarrow \infty} \frac{2.41 a \hbar^2 N - 1}{2\pi m} \frac{1}{V} = \frac{1.2 a \hbar^2}{m \pi l^3}.$$

As has been pointed out in § 3, the denominator of ϵ involves not precisely V , the volume of the space in which the particles are confined, but a "free volume" V' , say. An estimate of the difference yields the value $V - V' = 29.4\pi a^3$.

ACKNOWLEDGMENT.

The author is indebted to Professor N. F. Mott, of the H. H. Wills Physical Laboratory, Bristol, for suggesting this problem and the type of wave-function used in this paper.

REFERENCES.

- DINGLE, R. B., 1949, *Proc. Camb. Phil. Soc.*
 FOWLER, R. H., and JONES, H., 1938, *Proc. Camb. Phil. Soc.*, **34**, 573.
 HEITLER, W., 1927, *Zeit. f. Physik*, **44**, 161.
 LENZ, W., 1929, *Zeit. f. Physik*, **56**, 778.
 LONDON, F., 1936, *Proc. Roy. Soc.*, **153**, 576; 1938, *Phys. Rev.*, **54**, 947; 1939, *J. Phys. Chem.*, **43**, 49.
 MASSEY, H. S. W., and BUCKINGHAM, R. A., 1938, *Proc. Roy. Soc.*, **168**, 378.
 MOTT, N. F., 1949, *Phil. Mag.*, **40**, 61.

LII. *New Methods for the Numerical Solution of Algebraic Equations.*

By Professor A. PORTER, M.Sc., Ph.D. *, and C. MACK, M.A. ††

[Received January 24, 1949.]

I. INTRODUCTION.

IN the theoretical study of linear mechanical and electrical systems the problem of determining the roots of algebraic equations of the fourth or higher degree frequently arises. Often, more than a hundred equations may have to be solved in the study of one system alone. Some new methods of solving such equations are presented here which are particularly suited to the above type of problem where the coefficients of successive equations do not differ greatly.

For the fourth degree equations and for the fifth and sixth degree equations where the roots have real parts all of one sign (as in most practical problems) these methods are quicker than the root squaring method of Graeffe (1837) which is usually employed. Other methods have been suggested in recent years (see Lin 1941, Sharp 1941, Liu 1941) but are not always satisfactory.

They are, essentially, trial and error methods and have the following advantages over Graeffe's method: (i) a mistake made early in the process

* Military College of Science, Shrivenham.

† Shirley Institute for Cotton Research, Manchester.

‡ Communicated by the Authors.

does not affect later work, (ii) no modification of the procedure is required if complex roots are present, (iii) only the final stages need be worked to the accuracy required.

Three different (though in some respects similar) methods, referred to as Methods A, B and C, are described in this paper and were developed at the Radar Research and Development Establishment, Malvern, Wores., in connection with the design of automatic tracking radar equipment in 1942.

2. THE NUMERICAL SOLUTION OF QUARTIC EQUATIONS.

2.1. *Fundamental Relations.*

In both Methods A and B we factorize the given quartic into quadratic factors and then solve these separately. Given a real quartic polynomial it is known that it can be factorized into two real quadratic factors, thus

$$f(x) = x^4 + c_1x^3 + c_2x^2 + c_3x + c_4 = (x^2 + ax + b)(x^2 + lx + m). \quad (2.1.1)$$

Equating coefficients of powers of x we obtain

$$\left. \begin{array}{ll} \text{(i)} & a + l = c_1; \\ \text{(ii)} & b + al + m = c_2 \\ \text{(iii)} & bl + am = c_3; \\ \text{(iv)} & bm = c_4 \end{array} \right\} \quad (2.1.2)$$

2.2. *Method A.*

This consists of choosing a likely value a_r for a , solving equations 2.1.2 (i), (ii) and (iii) for the corresponding values of l , m and b (call them l_r , m_r and b_r) and then seeing how close $b_r m_r$ is to c_4 . We repeat with other values of a_r until $b_r m_r$ is as near c_4 as desired. We have thus found values of a and b which simultaneously solve 2.1.2 and which therefore factorize $f(x)$.

The formulæ for l_r , b_r and m_r in terms of a_r are

$$\left. \begin{array}{ll} \text{(i)} & l_r = c_1 - a_r \\ \text{(ii)} & b_r = [c_3 - a_r(c_2 - a_r l_r)] / (l_r - a_r) \\ \text{(iii)} & m_r = c_2 - a_r l_r - b_r \end{array} \right\} \quad (2.2.1)$$

It can be seen from (2.2.1) that $b_r m_r$ is a continuous function of a_r except at $a_r = c_1/2$, when b_r and m_r are both infinite but of opposite sign (hence $b_r m_r \rightarrow -\infty$ on both sides of $a_r = c_1/2$). Hence if a_1 and a_2 are such that $b_1 m_1 - c_4$ and $b_2 m_2 - c_4$ are of opposite sign there is a value a_0 between them such that $b_0 m_0 = c_4$. Once a fairly small value of $b_r m_r - c_4$ has been obtained linear interpolation in a_r from the values of $b_r m_r - c_4$ enables great accuracy to be obtained rapidly.

If all the roots are known to have real parts of the same sign then we know that there is one value of a between 0 and $c_1/2$. In this case the first two values of a_r considered are usually 0 and $c_1/4$.

2.3. *Application of Method A to a Numerical Case.*

We now solve $f(x)=x^4+2x^3+10x^2+5x+1=0$.

Routh's criteria (1859) show that all roots have negative real parts. Hence one value of a lies between 0 and 1. For our first two values in the following table we choose 0 and 0.5 and subsequent values are chosen by interpolation from the values of $b_r m_r - c_4$.

a_r	l_r	$(l_r - a_r)$	$[c_3 - a_r(c_2 - a_r l_r)]$	b_r	m_r	$b_r m_r - c_4$
0	2	2	5	2.5	7.5	17.75
0.5	1.5	1.0	0.375	0.375	8.875	2.328
0.6	1.4	0.8	-0.496	-0.62	9.780	-7.06
0.52	1.48	0.96	0.2002	0.2085	9.022	0.88
0.53	1.47	0.94	0.1129	0.1201	9.1008	0.09
0.531	1.469	0.938	0.10420	0.11109	9.10887	0.0119
0.53115	1.46885	0.93770	0.10289	0.10973	9.11009	-0.00035

Hence $f(x)=(x^2+0.53115x+0.10973)(x^2+1.46885x+9.11009)$. The right-hand side on multiplication is

$$x^4 + 2.00000x^3 + 10.00000x^2 + 5.00000x + 0.99965.$$

Equating the two quadratic factors to zero in turn we obtain as roots of $F(x)=0$,

$$-0.26558 \pm i0.19799; \quad -0.73443 \pm i2.9276.$$
2.4. *Method B.*

This can be applied to all cases irrespective of the signs of the roots. We again solve the equations (2.1.2) but this time start by choosing likely values for b instead of a . Given $b=b_r$, say, then, solving (2.1.2) (iv), (iii) and (i) for the corresponding values of m , a and l , we obtain

$$m_r = c_4/b_r, \quad a_r = (c_1 b_r - c_3)/(b_r - m_r), \quad l_r = c_1 - a_r. \quad (2.4.1)$$

Then we examine the value of $b_r + a_r l_r + m_r - c_2 = g_r$, say, for its closeness to zero.

From (2.1.2) (iv) it is clear that one value of b must satisfy

$$0 \leq |b| \leq \sqrt{|c_4|}. \quad (2.4.2)$$

Now consider the continuity of g_r . If $c_4 < 0$ then as $b_r \rightarrow \pm 0$, m_r and hence $g_r \rightarrow \mp \infty$. Further as $b_r \rightarrow \pm \sqrt{|c_4|}$, $g_r \rightarrow [c_1^2/4 + c_3^2/(4c_4) - c_2] = H$, say. Hence a value of b lies between 0 and $\pm \sqrt{|c_4|}$ according as $H \leq 0$. If $c_4 > 0$, $g_r \rightarrow -\infty$ as $b_r \rightarrow +\sqrt{c_4}$ while $g_r \rightarrow +\infty$ as $b_r \rightarrow +0$. Hence there is always a positive value of b between 0 and $\sqrt{c_4}$ in this case.

2.5. *Application of Method B to a Numerical Case.*

Consider the equation

$$F(x) = x^4 - 5x^3 + 7x^2 + 9x - 4 = 0. \quad (2.5.1)$$

Here $c_4 < 0$ and $H < 0$, hence the range $-\sqrt{|c_4|}$, 0 is to be explored. We shall take our first value b_1 of b_r as $-\frac{1}{2}\sqrt{|c_4|}$ and from the sign of g_r decide which side of this value to take b_2 .

b_r	m_r	$b_r - m_r$	$(c_1 b_r - c_3)$	a_r	l_r	g_r
-1	4	-5	-4	0.8	-5.8	-8.64
-0.5	8	-8.5	-6.5	0.765	-5.765	-3.91
-0.25	16.0	-16.25	-7.75	0.477	-5.477	6.135
-0.40	10.0	-10.40	-7.00	0.673	-5.673	-1.220
-0.365	10.959	-11.324	-7.175	0.6336	-5.6336	0.025
-0.36565	10.93942	-11.30507	-7.17175	0.634385	-5.634385	-0.00060

In obtaining the fifth value -0.365 of b_r we averaged the results obtained by linear interpolation from the third and fourth values of g_r and from the second and fourth values, since g_r is obviously not linear over the range involved. Hence

$$F(x) = (x^2 + 0.634885x - 0.36565)(x^2 - 5.634385x + 10.93942) \\ = x^4 - 5.00000x^3 + 6.99940x^2 + 9.00002x - 4.00000.$$

And the roots of $F(x) = 0$ are approximately

$$-1.00002, \quad +0.36564, \quad +2.81719 \pm i 1.73288.$$

2.6. Nomogrammatic Charts for Quartics (Mack and Tomlin 1945).

In these the solution of equations (2.1.2) is reduced to determining the intersection of a straight line and a family of curves. Assuming $c_4 > 0$ and writing $a = c_1 A$, $l = c_1 L$, $b = B\sqrt{c_4}$, $m = M\sqrt{c_4}$, (2.1.2) becomes

$$A + L = 1, \quad (B + M)\sqrt{c_4} + c^2 AL = c_2 \\ (BL + AM)c_1\sqrt{c_4} = c_3, \quad BM = 1.$$

Transforming to new variables $\alpha = AL$, $\beta = B + M$, these become, after manipulation,

$$(i) \beta + k_1 \alpha = k_2, \quad (ii) \alpha(\beta^2 - 4) = k_3 \beta - (1 + k_3^2),$$

where $k_1 = c_1^2/\sqrt{c_4}$, $k_2 = c_2/\sqrt{c_4}$, $k_3 = c_3/(c_1\sqrt{c_4})$. (i) represents a straight line in the α, β plane while (ii) represents a family of curves dependent on a single parameter k_3 . These curves have been plotted in such a manner that interpolation to three figures is simple. A similar set of curves for the case of negative c_4 's have been plotted and a single curve suffices for cubics. The time required to solve cubics and quartics to three-figures accuracy rarely exceeds five minutes.

3. APPLICATION OF METHOD A TO HIGHER DEGREE EQUATIONS.

3.1. Fundamental Relations.

Suppose $f(x) = x^n + c_1 x^{n-1} + c_2 x^{n-2} + \dots + c_n$. Let $(x^2 + ax + b)$ be a factor of $f(x)$. Then

$$f(x) = (x^2 + ax + b)(x^{n-2} + d_1 x^{n-3} + \dots + d_{n-2}). \quad (3.1.1)$$

Equating coefficients of powers of x we obtain

$$\begin{aligned} a+d_1=c_1, \quad b+ad_1+d_2=c_2, \quad bd_1+ad_2+d_3=c_3, \\ bd_2+ad_3+d_4=c_4, \quad \quad bd_{n-4}+ad_{n-3}+d_{n-2}=c_{n-2}, \\ bd_{n-3}+ad_{n-2}=c_{n-1}, \quad bd_{n-2}=c_n. \quad \quad (3.1.2) \end{aligned}$$

We assume a value a_r for a , then solve for $d_1, d_2, \dots d_{n-2}$ in terms of the known coefficients c_s , the assumed value a_r and b which is unknown as yet. Substituting in the last two equations of (3.1.2) we obtain two algebraic equations for b . When we have found a value a_0 for a_r such that these two equations have a common root b_0 , then a_0, b_0 are a possible pair of values for a, b in (3.1.2) and hence $x^2+a_0x+b_0$ is a factor of $f(x)$.

The procedure is fairly short for the fifth and sixth degrees, but becomes somewhat lengthy above that. Real roots can be found incidentally during the process.

3.2. General Procedure for Fifth and Sixth Degrees.

Given $f(x)=x^n+c_1x^{n-1}+c_2x^{n-2}+\dots+c_n=0$, where $n=5$ or 6 , proceed as follows. Assume a value a_r for a then calculate

$$\begin{aligned} L=c_1-a_r, \quad M=c_2-a_rL, \quad N=c_3-a_rM, \quad Q=c_4-a_rN, \\ L'=L-a_r, \quad M'=M-a_rL', \quad R=c_5-a_rQ, \quad S=c_6-a_rR, \\ L''=L'-a_r, \quad N'=N-a_rM'. \end{aligned}$$

When $n=5$, L'', N' and S can be omitted.

3.3. Fifth Degree.

The two equations in b are

$$(i) \ 0=Q-bM'+b^2 \quad (ii) \ 0=c_5-bN+b^2L'.$$

By eliminating the b^2 terms between (i) and (ii) we obtain

$$b=(L'Q-c_5)/(L'M'-N')=b',$$

say. When $b'(M'-b')=Q$, (i) is satisfied and therefore (i) and (ii) have a common root. Repeat with other values of a_r until $b'(M'-b')-Q$ is as near zero as desired.

A real root of $f(x)=0$ may be found by varying a_r until $R=0$ (for $a_r=a'$ say). Then $(x+a')$ is a factor of $f(x)$.

3.4. Sixth Degree.

The equations in b are

$$(i) \ 0=R-bN'+b^2L'', \quad (ii) \ 0=c_6-bQ+b^2M'-b^3. \quad (3.3.2)$$

Obviously $L''(ii)+b(i)$ reduces to a quadratic equation in b and eliminating the b^2 term between this and (i) a linear equation in b is obtained and so on.

This is best performed by the following procedure. Evaluate

$$\begin{aligned} T=N'/L'', \quad U=R/L'', \quad A=M'-T, \quad B=Q-U, \\ D=c_6-AU, \quad E=B-AT, \quad b'=D/E, \quad C=U-b'(T-b'). \end{aligned}$$

In general, when $C=0$, $x^2+a_r x+b'$ is a factor of $f(x)$. However, when $a_r=c_1/3$, $L''=0$, and C is also zero. In this case $b'=R/N'$ and only if this value of b' makes the r.h.s. of (3.3.2) (ii) zero is $(x^2+a_r x+b')$ a genuine factor*.

3.5. Choice of a_r .

When it is known that the roots have real parts all of one sign then for a $2m$ or $(2m+1)$ th degree equation $f(x)=0$ there must be one value a_0 of a_r such that $0 < ma_0 < c_1$. For, if a_1, a_2, \dots, a_m correspond to a set of m quadratic factors whose product equals $f(x)$ (multiplied by a linear factor if necessary) then $a_1+a_2+\dots+a_m \leq c_1$.

Hence 0 and $c_1/(2m)$ are the best first two values for a_r . Note that all values of b must be positive in this case.

When one quadratic factor has been obtained it may be divided into $f(x)$ and the resulting quotient factorized.

3.6. Numerical Case.

Applying Method A to the solution of the sextic,

$$f(x)=x^6+4x^5+16x^4+29x^3+31x^2+12x+2=0.$$

We obtain the following table:—

a_r	L''	m'	N'	Q	R	D	E	b'	C
0	4	16	29	31	12	-24.25	-35.44	0.685	-1.51
1	1	11	5	15	-3	20	-12	-1.667	8.33
0.5	2.5	12.75	15.5	20.062	1.969	-3.16	-21.33	0.1481	-0.108
0.55	2.35	12.507	14.365	19.316	1.376	-1.744	-20.357	0.0857	0.0692
0.530	2.41	12.603	14.815	19.608	1.608	-2.307	-20.744	0.1112	-0.0041

Linear interpolation in both a_r and b' from the last two values of C gives $(x^2+0.53112x+0.10978)$ as an approximate factor of $f(x)$. On division into $f(x)$ the quotient is $x^4+3.46888x^3+14.04783x^2+21.15811x+18.22033$ with a remainder $0.00008x-0.00023$. The quartic can now be solved.

3.7. Method B applied to Sixth Degree Equations.

If, in the set of general relations (3.1.2) we assume a value b_r for b then from the last $(n-2)$ equations we can evaluate d_1, d_2, \dots, d_{n-2} in terms of the c_s coefficients, b_r and a . Substituting in the first two equations we obtain two algebraic equations for a . When b_r is chosen so that they have a common root a' then $(x^2+a'x+br)$ is a factor of $(f(x))$.

For the sixth degree equation

$$f(x) \equiv x^6 + c_1 x^5 + c_2 x^4 + c_3 x^3 + c_4 x^2 + c_5 x + c_6 = 0.$$

* For a_r near $c_1/3$ see whether C/L'' is zero (instead of C) since $C/L'' \rightarrow$ a non-zero limit as $a_r \rightarrow c_1/3$ unless $c_1/3$ is a true value of a . Note also that, if the sign of C/L'' differs for any two values of a_r , a true value of a lies between them.

proceed as follows. Choose a value b_r , then evaluate,

$$\begin{aligned} q_r &= c_6/b_r, & A &= b_r - q_r/b_r, & B &= c_5/b_r - b_r c_1, \\ C &= q_r + b_r c_2 - c_4 - b_r^2, & D &= c_2 - 2b_r - q_r/b_r, & E &= b_r c_1 + c_5/b_r - c_3. \end{aligned}$$

The two equations for a are now

$$(i) \quad Aa^2 + Ba + C = 0, \quad (ii) \quad a^3 - c_1 a^2 + Da + E = 0. \quad (3.7.2)$$

Combining these two equations so that they reduce to a linear is most conveniently performed thus :—Evaluate

$$\begin{aligned} F &= B/A, & G &= C/A, & H &= c_1 + F, & K &= D - G, \\ a' &= -(HG + E)/(HF + K), & Z &= G + a'(F + a'). \end{aligned}$$

When $Z/A=0$ ($x^2 + a'x + b_r$) is a factor of $f(x)$. Note that as $b_r \rightarrow \sqrt[3]{c_6}$ both A and Z approach zero whatever $f(x)$. Also, when $b_r \rightarrow 0$, $Z \rightarrow 0$ but $Z/b_r \rightarrow -1$. If Z/A changes sign for two positive values of b_r there is a true value of b between them, and the same is true for two negative values.

The foregoing can be applied to any sixth degree equation whatever the signs of its roots for there must be one value of b such that

$$0 < |b| \leq \sqrt[3]{|c_6|}.$$

4. METHOD C.

4.1. Procedure.

Let $f(x) = x^n + c_1 x^{n-1} + \dots + c_n$. If $f(x)$ is divided by an arbitrary quadratic factor ($x^2 + ax + b$) we obtain a remainder $Fx + G$, where F and G are clearly functions of a and b . Now keep a fixed at a_1 , and let b take in turn the values b_1, b_2, \dots, b_p . Values of $F(a_1, b_r)$ are obtained and by interpolation a value b' can be obtained for which $F(a_1, b') = 0$. Similarly b'' is obtained such that $G(a_1, b'') = 0$. By repeating this process for other fixed values of a , points in the a, b plane can be found at which $F(a, b) = 0$, joining these points by a smooth curve we obtain approximately the curve $F(a, b) = 0$. Similarly the curve $G(a, b) = 0$ is obtained. The points of intersection of these two curves give possible values of (a, b) for which $(x^2 + ax + b)$ is a factor of $f(x)$. Investigation in their neighbourhood will improve the accuracy of a and b or, better, methods based on a generalized Newton Formula can be used (Bairstow 1914).

4.2. Limits for a and b .

When it is known that the roots all have negative real parts then we can state the following inequalities :—If the degree of $f(x) = 2m$ or $2m+1$ we have shown (3.5) that there is a value of a such that

$$0 < a \leq c_1/m. \quad (4.2.1)$$

Further, all values of a satisfy

$$0 < a < c_1. \quad (4.2.2.)$$

From the relations (3.1.2) using the fact that all the c_s and d_s are positive we deduce that, given a , the corresponding values of b satisfies.

$$c_n/c_{n-2} < b < c_2 - a(c_1 - 2), \quad . \quad . \quad . \quad . \quad . \quad (4.2.3)$$

and

$$c_{n-1} > c_n a/b > c_{n-1} b c_{n-3}. \quad . \quad . \quad . \quad . \quad . \quad (4.2.4)$$

When it is not known that the coefficients are all of one sign then an inequality can be stated for b when $n=2m$, namely

$$0 < |b| < |c_n|^{\frac{1}{m}}.$$

No simple inequality for a has been found except when it is known that all the roots are complex when $|a| < 2\sqrt{|b|}$.

REFERENCES.

- BAIRSTOW, 1914, *Aeronautical Research Council Reports and Memoranda*.
 GRAEFFE, 1837, *Autolsung der hoheren numerischen Gleichungen* (Zurich 1837).
 Method improved by Brodetsky and Smeal, 1924, *Proc. Camb. Phil. Soc.*,
 22, 83.
 LIN, S. N., 1941, *Method of successive approximation of evaluating the roots of cubic and higher equations*, *J. Maths. and Phys.*, Vol. 20, No. 3.
 LIU, Y. J., 1941, 'Servomechanisms'—charts for finding roots of third and fourth degree equations, *J. Mass. Inst. Tech.*
 MACK and TOMLIN, 1945, *Radar Research and Development Establishment Research Report No. 274*.
 ROUTH, 1859, *A treatise on the dynamics of rigid bodies* (Part II, p. 170), Macmillan, 6th ed. 1905.
 SHARP, 1941, *Comparison of methods for evaluating the complex roots of quartic equations*, *J. Maths. and Phys.*, Vol. 20, No. 3.

LIII. Heavy Splinters in Cosmic Ray Stars.

By Dr. A. BONETTI * and Miss C. DILWORTH;
 Centre de Physique Nucleaire de Bruxelles †.

[Received January 31, 1949.]

[Plates I. & II.]

It has long been recognized that some of the heavy splinters which are emitted in cosmic ray stars detected in photographic emulsions are not all of them alpha-particles but must include particles of higher mass and charge.

Apart from an event observed by Heitler, Powell and Fertel (1939) and the well-known lithium fragment leading to the hammer, it has been difficult to prove the existence of these heavy fragments. The question is certainly of interest in the study of the disintegration of nuclei by high energy cosmic rays.

* On leave of absence from the Department of Physics of the University of Genoa.

† Communicated by the Authors.

Recently Dr. Wilson presented at the Bristol Symposium evidence observed by M. G. Nooh and S. R. Haddara of a heavy fission like fragment track of charge between 4 and 12, photographed in Wilson cloud chamber at Jungfraujoch (3300 m.) (Nooh and Haddara). It was agreed then that no similar event had been detected in photographic plates at this or similar altitudes.

Heavy charged particles in cosmic rays have only been observed as single tracks of very high energy in high altitude flights (30,000 m.) (Freier, Lofgren, Ney, Oppenheimer, Bradt and Peters 1948).

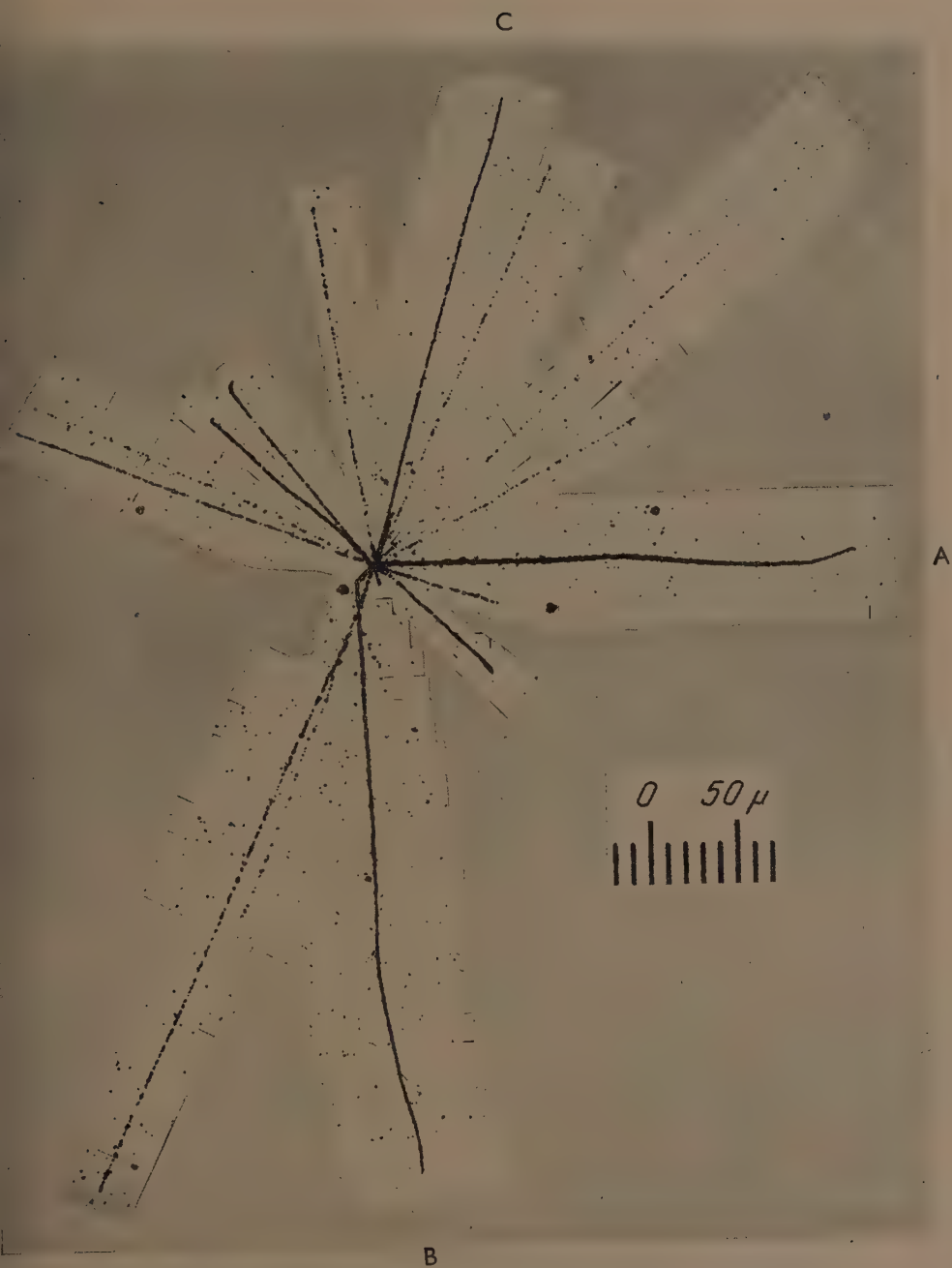
Owing to their high frequency of scattering, highly charged particles of great energy will be difficult to detect in a 100 micron emulsion. In an exposure at Pic du Midi (2850 m.), 300 micron Ilford C₂ plates were exposed in sandwich, so presenting an effective thickness of 600 micron. Even development was ensured by the temperature method. Under these conditions it was possible to establish the existence of highly charged particles in cosmic ray stars, though their occurrence with high energy is rare.

One of these events is shown in Pl. I.

This is a photograph of a star of 20 branches, of which 10 are certainly protons, 4 are so short that it is impossible to determine their nature, 2 are alpha particles of low energy, 1 is possibly a deuteron. Of the 4 heavy particles of long range, C, D are alpha particles; A is 286 micron long, shows δ rays of three grains in length, *i. e.* 20 keV. energy, up to 100 micron from the end of its range, and has the characteristic tapering of a fission fragment at the end. B, which is 330 micron long and suffers a collision 15 micron from the centre of the star, shows no δ rays but is more heavily ionizing than the alpha particle C of the same dip. These points are illustrated in Pl. II., in which the beginning and end of both heavily ionizing fragments are compared with those of particle C.

The apparently greater ionization of the particle B before the collision is due to the presence of a short track lying above it in the emulsion. This can be seen when the plate is turned over and viewed through the backing glass, which it was possible to do with a $\times 41 \times 40$ dry objective since the plates were poured on thin glass (1 mm.).

To estimate the charge of the particle A we used the method of Bradt and Peters (1949). The minimum velocity at which δ rays of a given energy are produced can be calculated from the formula of Mott (1929). The charge can then be deduced from the residual range R of the particle at this velocity by transforming the range-energy relation for alpha particles, using the relation $R = (M/Z^2)f(v)$, where M is the mass, Z is the charge of the particle and assuming that $M = 2Z$, as is most likely. The value so obtained was $Z = 10$. With this charge and a total range of 286 micron, the fragment must have been emitted with an energy of about 260 MeV. The above range velocity relation is not strictly valid for highly charged particles due to capture of electrons at the end of the range, and we therefore tend to underestimate the value of the charge.



Cosmic Ray Star with Emission of Heavy Fragments of Long Range.



Comparison of the density of track at origin and end of alpha particle (c) with those of heavy fragments (A) and (B).

As a check on the reliability of the method we repeated the calculation for the Li_3^8 branch of a hammer track found in a star in the companion plate of the sandwich. This branch is 500 micron long and shows δ rays of 20 keV. for the first 20 micron from the centre of the star. The value of the charge found was $Z=3.3$.

The charge of the particle B we estimated by comparing the breadth of tracks A, B and C, and found it to be about 4. The energy of this particle, which has a range of 330 micron, should then be of the order of 90 MeV.

It may be of interest to give also an estimate of the total energy of the star. The total charge visible in the disintegration is 35 ± 6 . By comparing the grain densities of the tracks which end in the emulsion with those which do not, we found approximately the energy carried by each particle. The sum of these is 550 ± 50 MeV. The total energy required to give the observed disintegration, when we add the energy necessary for the liberation of the 20 particles, is then about 650 MeV. This quantity gives the minimum energy of the incident radiation since we have not taken into account the energy represented by neutrons and high energy electrons and protons, which are invisible in this emulsion. We can assume that the neutrons carry a total energy at least equal to that of the protons, *i.e.* 150 MeV.; then if the disintegrated nucleus is one of Bromine, the total mass defect is approximately 300 MeV. and the total energy becomes of the order of 1000 MeV.

We wish to express our thanks to Prof. Cosyns for extending to us the facilities of his laboratory at the Centre de Physique Nucléaire de Bruxelles; to Dr. Occhialini for suggesting this work and his help and encouragement in carrying it out. We thank also our colleagues of this laboratory for their kind assistance and Prof. Schönberg for helpful discussions. We are indebted to Mr. Waller of Ilford Ltd. for the special emulsions and to the members of the Observatoire du Pic du Midi for the exposure. One of us (A. B.) was enabled to carry out this work here through the kindness of the Physical Institute of the University of Genoa.

Note added in proof:—Other examples of the emission from stars of highly charged particles of long range have been reported in a recent publication by Hodgson and Perkins (1949).

REFERENCES.

- BRADT and PETERS, 1949 (in press).
FREIER, LOFGREN, NEY, OPPENHEIMER, BRADT and PETERS, 1948, *Phys. Rev.*, **14**, 213.
HEITLER, POWELL, and FERTÉL, 1939, *Nature, Lond.*, **144**, 283.
HODGSON and PERKINS, 1949, *Nature, Lond.*, **163**, 439.
MOTT, 1929, *Proc. Roy. Soc.*, **124**, 425.
NOOH and HADDARA (unpublished work).

LIV. *Notices of New Books and Periodicals received.*

Advances in Electronics. Edited by L. MARTON. (New York: Academic Press, Inc.) Price \$9.00.

THIS is a thoroughly useful volume, the first of a series, which the editor hopes to bring out annually. The aim is to give up-to-date summaries of various branches of the subject. Those included in the first volume are:—Oxide Coated Cathodes, by A. S. Eisenstein; Secondary Electron Emission, by K. G. McKay; Television Pick-up Tubes and the Problem of Vision, by A. Rose; The Deflection of Beams of Charged Particles, by R. G. E. Hutter; Modern Mass Spectroscopy, by M. G. Inghram; Particle Accelerators, by M. Stanley Livingston; Ionospheric Research, by A. G. McNish; Cosmic Radio Noise, by J. W. Herbstreit; Propagation in the FM Broadcast Band, by K. A. Norton; and Electronic Aids to Navigation, by J. A. Pierce.

Elasticity and Anelasticity of Metals. By C. ZENER. (University of Chicago Press.) Price 22s. 6d.

THIS book is mainly about anelasticity; that is to say, it treats such subjects as creep recovery, damping of vibrations and internal friction. The subject is one that Dr. Zener has made particularly his own, and it is largely due to him that so many of the mechanisms responsible for anelasticity are now understood. Some of these discussed in the book are:—Thermal currents, both transverse and intercrystalline, movements of interstitial atoms such as dissolved carbon, grain boundary flow and movement of dislocations. The book is a record of an important achievement in research and is warmly to be welcomed.

Crystals and X-Rays. By KATHLEEN LONSDALE. (G. Bell & Sons, Ltd.) Price 21s. 0d.

THIS work is written essentially for industrial administrators and technicians who might wish to gain a deeper understanding of the principles and applications of X-ray crystallography. It is a general survey written in simple language, and covers most of the more recent developments in this science.

The author's lucid style and helpful analogies maintain interest, while the 138 figures and 80 excellent reproductions of X-ray diffraction photographs should be of value even to the expert. Although the book is not essentially a work of reference, it is well indexed, and the limited selection of references to original works are well chosen.

Chamber's Six-figure Mathematical Tables. By L. J. COMRIE, M.A., Ph.D.
Vol. 1. Logarithmic Values. 42s. net. Vol. 2. Natural Values. 42s. net.

THE publication of these excellent and beautifully printed tables by the director of the Scientific Computing Services Ltd. will be widely welcomed. The functions tabulated include, as well as logarithms and trigonometrical functions, sinh, cosh, tanh and coth, exponential functions, powers and roots, the gamma function and various conversion tables. The book contains also tables of constants and a bibliography. The typography is all that can be desired, and as each volume contains nearly 600 pages the price seems very moderate.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

LV. *The Production of Cosmic Ray Stars.*

By Sir GEORGE THOMSON,
Imperial College, South Kensington*.

[Received February 28, 1949.]

SUMMARY.

A study is made of the present knowledge of the production of stars and single tracks by cosmic rays in photographic plates. In particular the deductions which can be made from the observed exponential variation of these effects and the constancy with varying height of the size of the star are examined; these are shown to be consistent with the view that one primary particle makes several effective collisions in the atmosphere. While no unambiguous decision can be made as to the mechanism of production some possibilities can be excluded. It is most likely that the effects are due *either* to nucleons passing down through the atmosphere and alternating in state between protons and neutrons as a result of nuclear collisions *or* to neutrons. In either case each "primary" particle must make several tracks or stars. The production of tracks by the primaries is direct, that of the stars may be direct or through a non-ionizing link, presumably some kind of neutral meson.

1. MORE information on the occurrence of cosmic ray stars has become available in the last few months, and it seems useful to consider what deductions can be made as to the production of these stars, relying as little as possible on theories of nuclear interaction.

2. The following facts seem well established for stars at moderate heights:

(1) The number of stars produced in a given kind of photographic plate diminishes exponentially with the amount of the atmosphere above it. The intensity falls by a factor e for about 160 gr./cm.² (Stetter and Wambacher 1939, Perkins 1947).

(2) A similar exponential diminution occurs if the plates are under ice or lead (Perkins 1948, George 1948 a and b).

(3) The transition effect if it exists at all, which is doubtful, is not marked (Bernadini *et al* 1947, George, private communication).

* Communicated by the Author.

(4) The absorption under ice and lead compared with that under air is roughly in the ratio of $A^{\frac{1}{2}}$ when reckoned per atom of absorber. A is the atomic weight (George 1948a, Harding *et al* 1949 a).

(5) Accompanying the stars are single tracks, mostly due to protons, whose number is proportional to the number of stars and varies in the same way as the number of stars for absorption in air, ice or lead. (Perkins 1947, Lattimore 1949.)

(6) The large majority of the stars are due to a non-ionizing radiation, though about 1 in 30 are due to slow mesons and a very few are the result of penetrating ionizing particles. The latter stars are in principle indistinguishable from penetrating showers but are very seldom seen in photographic plates, partly because of their intrinsic rarity and partly because most plates do not detect very fast particles. Stars are not associated with cascade showers and are therefore not due to any great extent to photons (Hazen 1944, W. M. Powell, 1946).

(7) While some of the "single tracks" must be due to primary protons, these are too few to account for more than 1 per cent. Tracks from stars outside the emulsion account for about 10 per cent, and the neutrons which presumably accompany the stars might be expected to produce about 1 per cent by knocking-on the protons in the emulsion. The remaining 88 per cent must have some independent cause. There are about $4\frac{1}{2}$ times as many "single tracks" as stars produced per c.c. per sec. (Lattimore, in publication).

(8) The proportion of stars having various numbers of tracks, and the average energy of these tracks for a star of given size does not vary appreciably with altitude or absorber if the composition of the emulsion remains the same (Bernadini, Cortini and Manfredini 1948, George 1948 b, unpublished results at Imperial College).

(9) Experiments on plates having layers of gelatine not containing AgBr show that the chance of the production of a star in an atom of C, N, O on the one hand, and Ag or Br on the other, goes roughly as $A^{\frac{1}{2}}$, where A is the weighted mean of each of these two groups of atoms (Harding 1949 b).

(10) The momentum given to the residue of the nucleus, which shows as a short "recoil" track, is markedly less than would be required if the stars were produced by a neutron striking the nucleus and coming to rest in the residue (Harding 1949 c).

(11) The distribution of stars in a plate is not uniform, but there is a distinct tendency for stars to occur within a distance of 2 mm. or less from each other. It is not at present possible to say if this is a general property of the star distribution, or the result of superposing a small proportion of close pairs on an otherwise random distribution (Leprince-Ringuet and Heidmann, Li and Perkins 1948).

(12) There is no difference in the way in which the energy is distributed among the tracks between stars due to slow negative mesons and those due to non-ionizing radiation. This applies not only to the majority of the tracks which can be explained as due to thermal evaporation from an excited nucleus, but also to the few long tracks which cannot be so explained (Perkins 1949).

(13) The number of slow neutrons in the atmosphere varies with height in the same way as the number of stars. The absolute number produced per c.c. per sec. is rather uncertain, but the latest measurements indicate that there are only 1.5 per star, which is, if anything, less than one would expect to accompany the charged particles visible in the stars (Yuan 1948: *see also* Korff and Cobas 1948, Simpson, 1948).

(14) The single tracks are fairly strongly directed in a vertical direction (Lattimore, in publication). The tracks in stars are sensibly isotropic.

3. Many of the above facts can be explained by supposing that the stars are produced by neutrons coming from the top of the atmosphere, assuming a cross-section for star production proportional to the geometrical cross-section of the nucleus, each neutron making one star. I wish to suggest in what follows that a somewhat different interpretation is equally possible and has indeed certain advantages.

4. Instead of supposing that each primary particle, which we will call a neutron, makes only one star, I shall suppose that it loses its energy in a number of steps, large enough to allow us to treat the loss as continuous. I shall consider later how far this extreme assumption resembles the truth. It will also be assumed that the energy lost appears in the near vicinity in the form of the stars and single tracks observed, and perhaps also of mesons, but the mechanism by which this is done will be left open for the moment.

5. Let $N(E, x)dE$ be the number of neutrons passing downward through one cm.^2 at a depth $x \text{ gr./cm.}^2$ below the top of the atmosphere in the energy range E to $E+dE$. Assume that the radiation is all vertical. Let $M(E)$ be the energy absorbed per neutron in passing through 1 gr./cm.^2 . In general the energy spectrum will change with depth, but there are particular cases in which it does not, and these will turn out to be the important ones for our purpose.

The $N(E, x)dE$ neutrons after going a further distance dx will have energies in the range E' to $E'+dE'$, where $E'=E-M(E)dx$.

Therefore
$$dE'=dE\left(1-\frac{dM}{dE}dx\right).$$

Accordingly

$$N(E', x+dx)dE'=N(E, x)dE,$$

or

$$\left(N - \frac{\partial N}{\partial E} M dx + \frac{\partial N}{\partial x} dx\right) \left(1 - \frac{dM}{dE} \cdot dx\right) = N,$$

whence

$$\frac{\partial N}{\partial x} - \frac{\partial(MN)}{\partial E} = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The solution of this equation is

$$MN = f\left(x + \int \frac{dE}{M}\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where f is an arbitrary function.

Now the energy spectrum of the stars and the proportion of "single tracks" to stars is independent of height*. This suggests that the energy spectrum of the radiation causing them is also independent of height.

Assuming that this is so, we have

$$N = \psi(x)\phi(E), \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where ψ and ϕ are functions to be determined if we can.

Substituting in (1) we have

$$\psi'(x)\phi(E) - \psi(x) \frac{d}{dE} \{M\phi(E)\} = 0,$$

where $\psi'(x)$ is the differential coefficient of $\psi(x)$. Hence

$$\frac{\psi'(x)}{\psi(x)} = \frac{1}{\phi(E)} \frac{d}{dE} \{M\phi(E)\};$$

since the left-hand side is a function of x only and the right-hand side of E only, each must equal a constant $-\alpha$, say. Hence

$$\psi'(x) = -\alpha\psi(x) \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and

$$\frac{d}{dE} \{M\phi(E)\} = -\alpha\phi(E). \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

From (4) we have at once

$$N = \exp(-\alpha x)\phi(E), \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

so that the observed exponential law is verified. This is strong support for the assumption of equation (3).

From (5) we have

$$M = -\frac{\alpha}{\phi(E)} \int \phi(E) dE = \frac{-\alpha}{N} \int N dE. \quad . \quad . \quad . \quad . \quad . \quad (7)$$

* This ceases to be true at great heights such as can be reached by balloon flights.

This shows that any distribution $\phi(E)$ requires, and can only be maintained by, the right variation of the loss of energy with the energy of the particle. However, the form of M is not unique since the integral in (7) contains an arbitrary constant.

Conversely, a particular form of M is associated with a particular energy distribution. Only if this energy distribution is initially present at $x=0$ can assumption (3) hold at all depths.

We have in fact from (5)

$$\frac{1}{M} \frac{dM}{dE} + \frac{1}{\phi(E)} \cdot \frac{d\phi(E)}{dE} = -\frac{\alpha}{M}.$$

Thus

$$\log(M) + \log \phi(E) = -\alpha \int \frac{dE}{M}$$

or

$$\phi(E) = \frac{1}{M} \exp \left[-\alpha \int \frac{dE}{M} \right]. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The arbitrary constant in the indefinite integral appears as an arbitrary factor in ϕ , and hence also in N , the number of primary particles.

We have not so far considered the boundary conditions and must now do so. There will be a lower limit A to the energy below which no stars will be formed. Neutrons whose energy falls below this limit pass out of observation and need not be considered further, we can if we like take ϕ or M as zero below this limit if necessary for convergence. The upper limit of energy is more important. Our view envisages neutrons steadily losing energy as they go downwards, those in each energy group being replaced by others originally of greater energy. Let us take an upper limit \bar{E} for the energy we consider. Then at each height we must supply enough neutrons of energy \bar{E} to make up the loss. In distance dx each neutron loses energy Mdx , hence the neutrons in a band of energy near \bar{E} of width $dE=Mdx$ have to be replaced. Their number is $N(\bar{E}, x)dE=N(\bar{E}, x)Mdx$, and this is the number that must be supplied. It will in general vary with height as $\exp(-\alpha x)$. In practice, however, we suppose that $N \rightarrow 0$ as $E \rightarrow \infty$, and that the inflow of infinitely fast neutrons is zero. Hence $MN \rightarrow 0$ as $E \rightarrow \infty$, or what comes to the same thing, $M \cdot \phi \rightarrow 0$ as $E \rightarrow \infty$.

This boundary condition has two effects; from (8) we see that it limits the form of M , since $\exp \left[-\alpha \int \frac{dE}{M} \right]$ must tend to zero as $E \rightarrow \infty$. Secondly it determines the constant of integration in (7). In fact we can replace the indefinite integral by a definite one and write

$$M\phi = \alpha \int_A^\infty \phi dE. \quad . \quad . \quad . \quad . \quad . \quad (9)$$

With these conditions equations (7) and (8) are always compatible, and if M and ϕ are finite at $E=A$, as we shall assume, $\int_A^\infty \phi dE$ is finite, and therefore also $\int_A^\infty N dE$, as it needs to be for obvious reasons.

A few special cases may be considered.

If $M = kE^S$, where $S \neq 1$, we have

$$\exp \left[-\alpha \int \frac{dE}{M} \right] = C \exp \left\{ \frac{\alpha}{k(S-1)E^{S-1}} \right\}.$$

This tends to zero if $S < 1$, but is finite if $S > 1$, which is therefore inadmissible.

$$\phi = \frac{C}{kE^S} \exp \frac{\alpha}{k(S-1)E^{S-1}}. \quad . \quad . \quad . \quad . \quad . \quad (10)$$

$$\text{If} \quad M = KE, \quad \exp \left[-\alpha \int \frac{dE}{M} \right] = CE^{-\alpha/k},$$

which tends to zero at infinity, so the law satisfies our criterion,

$$\phi = \frac{C}{k} E^{-\alpha/k-1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

thus ϕ , and so N , varies as a power of E .

In fact, if $N \propto E^{-q}$, we have

$$\alpha = k(q-1). \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

This is the only law for M which allows N to vary as a power of E , as can readily be seen by substituting E^{-q} for N in (7).

If $\phi = B \exp(-E)$, we find from (9), $M = \alpha$. This satisfies the condition that $\exp \left[-\alpha \int \frac{dE}{M} \right] \rightarrow 0$ as $E \rightarrow \infty$.

7. The most important feature of all these solutions is that N is the product of two factors, one a factor of x only, in fact an exponential, the other a function of E only, while the energy spectrum is the same at all heights. When the radiation passes from one medium to another the exponential law will be retained if M remains the same, apart from a constant factor, on passing from one medium to another. Otherwise there will be a transition layer in which the energy spectrum of the rays will change. Since in fact little or no transition is observed, we must suppose M has the same form for all nuclei. It should be remembered that M covers all forms of energy loss corresponding both to star and to track production and to any creation of mesons which may take place.

8. If the radiation is not vertical the above analysis should be modified. Rossi (1948) has shown how an isotropic radiation homogeneous as regards absorption varies with depth below the top of the atmosphere. If the component in each direction is exponentially absorbed the initial absorption will be faster than this exponential would indicate, but will approach it as the depth increases. In our case the fact that the tracks are so strongly directed downwards indicates at least as great—probably a greater—anisotropy in the radiation causing them. It seems likely therefore that we are in the region where a truly homogeneous radiation would show its natural rate of absorption. In other words, where the

remaining obliquity can be neglected. Since in any case the radiation with which we are concerned is not primary in the strict sense of the term, we have no means of knowing what its original distribution in direction was at the height where it is formed. Rossi's assumption of initial isotropy is the extreme case. It seems fairly safe to regard any modification of our results needed for this reason as a small effect.

9. The close connection between stars and single tracks requires either (a) they are both due directly to the same cause, presumably neutrons; (b) the stars cause the tracks through a non-ionizing link; (c) the tracks cause the stars through a non-ionizing link; (d) both stars and tracks are caused by the neutrons through non-ionizing links.

Since the tracks are strongly collimated and the stars not, it seems safe to exclude (b). It is harder to distinguish between (a) and (c), and at present it is best to leave the question open, treating the two possibilities separately. Possibility (d) seems much less likely, since a collision producing a new non-ionizing particle would be expected to leave some trace on the photographic plate and no type of track is known which can be assigned to this cause.

10. Stars can only be seen clearly when produced *either* in an emulsion or in the gas of a cloud chamber. In practice, stars produced in the wall of the chamber are usually only visible if they are of the rare collimated type which may be quite different in origin from the type we see in the plates.

If the stars are produced indirectly, we must distinguish between the absorption of the primary star-producing radiation in the matter surrounding the plate and the actual production of a star by the absorption of the secondary radiation in the nucleus of some atom in the plate or gas. It is possible, though not certain, that each secondary particle will usually produce only one star. We shall assume so to begin with.

If the stars are produced directly it is likely, from Harding's work (1947c) that each primary particle produces more than one star; it will probably also produce several single tracks. We must still distinguish between the absorption of the primary radiation as a whole and that part of the absorption of energy which is used in star production. The two may vary in quite different ways with the nature of the medium, cf. the production of fission by a beam of fast neutrons passing through matter containing variable amounts of uranium.

Let M_z , a function of E , be the energy loss per primary particle in traversing 1 gr./cm.² of material Z . In the case of direct production, let P'_z be the production of stars in atom species Z' per primary particle per atom per cc. It will, of course, be a function of E .

In the case of indirect production let M'_z be that part of M_z concerned with the secondary rays, let a be the range of those rays, assumed each of energy ϵ , in gr./cm., and let Q'_z be the production of stars in atom

species Z' per atom per c.c. in path of a secondary ray. It may or may not depend on the energy of the primary from which the secondary was formed.

Suppose some plates are placed in a cavity in an extended medium (which might be air). Then for *direct production* the number of stars per c.c. of emulsion per sec. is

$$\Sigma \int_A^{\infty} N P_z C_{z'} dE,$$

where $C_{z'}$ is the concentration of atoms Z' in the emulsion in atoms/c.c., and A is a lower limit of energy below which star production ceases. Assuming that the primary rays lose energy gradually, the variation of N with x is determined by M and has been examined in the previous section. We have seen that N is of the form $e^{-\alpha x} \phi(E)$, where $\phi(E)$ does not contain x nor α contain E . A law of this kind means that there is no transition effect on passing from one medium to another, provided M is the same function of E in both media. In such a case N will be the same function of E in both media, and the production of stars in similar plates placed in cavities in the two media will diminish exponentially with depth as N does, since $P_z, C_{z'}$ are the same. This is the case usually considered. If the plate is changed, the production of stars will, of course, change also, but provided N is of the form $e^{-\alpha x} \phi(E)$ the factor by which it changes will be the same at all depths. The relative number of stars produced in different atomic species, allowing for concentration, is now the ratios of the quantities P_z , *not* the quantities M_z which would determine absorption in continuous media of the different atomic species.

For *indirect production* we must distinguish various cases:—

(1) Let a be large compared with the thickness of packing, glass, etc., round the emulsion. Then the production of stars $*$ is

$$\sim \Sigma \rho \frac{a}{4} \int_A^{\infty} N \frac{M'_z}{\epsilon} Q_{z'} C_{z'} dE,$$

where the density ρ will be the average for a place within a distance a of the plate, and Z is the atomic number there. In this case the relative production of stars in different kinds of nuclei in the plate goes as $Q_{z'} C_{z'}$, assuming $Q_{z'}$ independent of the *primary* energy E .

(2) Let a be less than the thickness of the emulsion. Then star production in atom species Z' is

$$\rho \frac{a}{4} \int N \frac{\overline{M}'_z}{\epsilon} Q_{z'} C_{z'} dE,$$

* The number of secondary rays which start per sq. cm. in a layer 1 gr./cm.² is $N M'_z / \epsilon$. If the range of each is a the total length of track in a volume V , large compared with a^3 , is $\rho V N M'_z a / \epsilon$, where ρ is density. If n tracks cross a square cm. placed at random, the total length of track in a small cube side s is $\sim 4ns^3$. Hence track length per c.c. is $\sim 4n$, therefore $n \sim \rho \frac{N}{4} \frac{M'_z}{\epsilon} \cdot a$.

where M'_z is replaced by M' , which is defined as the weighted mean of M' taken over the atoms of the emulsion, and ρ , a , refer to the emulsion. Here again the proportion of stars from different kinds of nuclei goes as $Q_z C_z$.

(3) a intermediate between these limits. The secondary radiation will depend on the nature of the material round the plate. As this changes, the star production should vary, unless the production of secondary radiation in a medium is always proportional to its absorption, but the proportion between different kinds of nuclei is unchanged.

(N.B.—The problem of *tracks* is different, for single tracks are observed in a plate, though produced elsewhere. The thickness for the "transition effect" in this case measures the range of the tracks.)

It will be seen that either assumption can be made to account for the experimental facts of §2, except that a connecting link seems necessary to account for the close pairs of stars (§2 (11).) Since, however, these may only represent a minority of stars, the argument is not a strong one as regards the majority.

11. It is now necessary to see how far the initial assumption of continuous energy loss by the neutrons is a reasonable approximation to the truth. As a preliminary to this, it is desirable to consider the single tracks in a little more detail.

12. There are several not improbable methods by which tracks might be produced:—(a) A single nucleon might pass through many nuclei changing its state between neutron and proton. The tracks seen in the plate could then be the last stage of this nucleon with its energy reduced to 30 MeV or under. (b) The neutron might remain a neutron but knock out a proton with moderate energy, perhaps not much exceeding 30 MeV. If the original neutron has an energy much exceeding this, it may be expected to repeat the process several times. (c) As in (b), but the neutron having less energy initially either remains in the nucleus, or (d) emerges with too little energy to give another track. The experimental evidence enabling us to distinguish between these possibilities is not so great as one would like. If one were certain that the number of neutrons in the atmosphere is as low as appears from the work of Yuan*, one could exclude (d) because it requires a slow neutron for each track, and since there are $4\frac{1}{2}$ single tracks per star this would double the production of neutrons, already uncomfortably large. Process (b) would give a smaller number which might be admissible. If more observations of the start of single tracks were available, it might be possible to observe the tracks due to the recoil of the residue of the nucleus, and by determining its momentum get some discrimination

* *Note added in Proof.*—The work of Toby (*Phys. Rev.*, **75**, p. 894, 1949) requires a considerably larger production of neutrons of the order of 6 per star at sea-level. The matter is still most uncertain.

between the above hypotheses. Another relevant fact is that protons of from 4×10^8 to 10^9 momentum units (MeV/c) have been observed among the cosmic rays. Rossi's (1948) estimate from the data available is that about 165 particles/cm.²/day occur in this momentum range (energy $8 \times 10^7 - 4.2 \times 10^8$ ev.) at the height of the Jungfrauoch. At the same height about 16 tracks/cm.²/day are found in the photographic plates. The apparent absorption of the cosmic ray protons is about 125 gr./cm.², according to Rossi, but this seems a very rough estimate, it might not be very different from that of the tracks, ~ 160 gr./cm.²

On view (a) we should regard the tracks as the ends of the cosmic ray protons when the energy had fallen below 30 MeV. We should expect a considerable number of changes of energy to take place in the life of a particle and the analysis of §5 should apply. From the known loss of energy of a proton, it would be possible, assuming exponential absorption, to calculate the relative numbers in the energy range 0–30 MeV and 80–420 MeV, but we do not know what energy loss, if any, occurs in the collisions which result in the change of state or what is the fraction of the time in either state. If *no* energy loss is assumed and about equal time in the two states, the ratio of the number of tracks in the two energy regions would be of the order 100 : 1 instead of 165 : 16, but the assumption of energy loss would alter the calculated ratio in the right direction.

Alternatives (c) and (d) imply that the neutrons making the tracks have much less energy than those making the stars. It seems unlikely that they are an entirely separate class because this would leave the close proportionality between tracks and stars unexplained. Since there are $4\frac{1}{2}$ times as many tracks as stars they cannot each mark the end of a star-producing neutron. It seems likely then that either (a) or (b) is correct with the evidence of the number of neutrons in the atmosphere slightly favouring (a). Either of these views implies that a primary particle will undergo energy losses in stages, apart altogether from any production of stars it may cause. In either case star production may be direct or indirect. Thus the collision which, in case (a), usually leads to a change of state, may alternatively produce a star, or it may produce some sort of neutral meson which in turn produces a star. It seems curious that the particle should not sometimes produce a star when in its proton state. That it so seldom does so is an argument against (a). In case (b) the collision that produces a proton may exceptionally (1 in $5\frac{1}{2}$ times) give rise to a star. If the star production is direct we know that there is generally enough energy left over for other stars, though this energy may, of course, be dissipated in single tracks. Thus in case (b) with direct star production, we may expect to have about a dozen energy losses in all, enough to make it reasonable to apply a continuous theory. If the star production is indirect, there may seldom be more than one such act in the life of a primary (though there must sometimes at least be more than one "link" produced at a time to explain the close pairs of stars). In this case the total of separate acts need not be so large,

but would still be over five on an average. The theory would probably be qualitatively correct but no more. In case (a) the main energy loss will be continuous for perhaps half the time, and the conditions should be satisfied.

13. The above theory, or rather theories, for there are several alternatives, appear to account for the known facts. We must suppose that M_z is proportional to α and so to $A^{-\frac{1}{3}}$, since M measures the loss per gr./cm.² In the case of direct production of stars P_z must go as $A^{\frac{1}{3}}$ at least for elements up to silver. For indirect production Q_z must follow this law if the range of the connecting-link is greater than the thickness of material round the emulsion, and M'_z/M_z must be independent of the atomic species. If a is small compared with the thickness of glass, paper, etc., round the emulsion, and this is constant for all the plates used, the relation of M'_z to M_z does not matter. The plate with its packing acts as a detector of the *primary* radiation and the proportion of secondary in the surrounding matter is irrelevant.

14. There is an objection which may at first sight be made against these or any other theories which require a single particle to make many collisions. Taking the nuclear cross-section as $\pi(1.5)^2 A^{\frac{1}{3}} \times 10^{-26}$, a particle will be expected to make one collision for every 56 gr./cm.² of air traversed. Now the observed penetration, $1/\alpha$, is 160 gr./cm.² which would seem to allow only 3 collisions per particle, not a large number. This objection is invalid for the following reason: corresponds not to the disappearance by absorption of individual particles but to the net decrease in the number of particles in a given energy range caused by the transfer of particles from one energy group to another. This net decrease depends on the energy spectrum of the particles. The depth from the region where the neutrons may be supposed generated to sea level is at least 900 gr./cm.², and in this distance we could expect $\frac{900}{56} = 16$ collisions, though one could not say *a priori* what proportion of these would result in any kind of nuclear change.

15. The difficulty of the other theory of constant cross-section and once-and-for-all impact is that it conflicts with the experiment on the motion of the residual nucleus. Further, in order to explain the observed exponential absorption it would be necessary to suppose that the cross-section is independent of the energy of the neutrons. This would be reasonable if the cross-section required were approximately equal to the geometrical cross-section of the nucleus—every hit would be effective. The figures of the paragraphs above show that it comes out at $\frac{1}{3}$; it seems rather unlikely that this fraction should not depend on the energy of the incident particle. If it is modified to include extra production of tracks or other stars by a particle which has already made one star, it will predict a change in the energy distribution and number of the particles with height, and will not predict exponential absorption except for very special distributions of the original energy spectrum.

Such distributions will gradually merge into those calculated above, as the number of separate acts involving loss of energy is supposed to increase.

Note.—While the above was in manuscript I had the opportunity, by the courtesy of Professor Janossy, of seeing a Paper by Dr. Heitler and himself, now in course of publication. This paper deals with a problem closely connected with that contained here and leads to similar conclusions. It is more detailed in that it considers the probability fluctuations caused by cosmic ray particles penetrating various thicknesses of nuclear matter during their passage through the atmosphere. On the other hand, the Law of Energy Loss considered is more restricted, and is, in fact, the only one consistent with a Power Law for the distribution of energy in the primary rays. It is also primarily concerned with the production of penetrating showers rather than stars, and for these reasons it seems desirable to publish the present paper.

I wish to thank Dr. Barford for valuable discussions on some of the mathematical points in this paper.

REFERENCES.

- BERNADINI, G., CORTINI, G., and MANFREDINI, A., 1948, *Phys. Rev.*, **74**, 845.
 GEORGE, E. D., 1948 a, *Nature, Lond.*, **162**, 333.
 GEORGE, E. D., 1948 b, *Bristol Symposium*.
 HARDING, J. B., LATTIMORE, S., LI, T. T., and PERKINS, D. H., 1949 a, *Nature, Lond.*, **163**, 319.
 HARDING, J. B., 1949 b, *Nature, Lond.* (in publication).
 HARDING, J. B., 1949 c, *Phil. Mag.*, **40**, 540.
 HAZEN, W. E., 1944, *Phys. Rev.*, **65**, 67.
 KORFF, S. A., and COBAS, A., 1948, *Phys. Rev.*, **73**, 1010.
 LATTIMORE, S., 1949, *Phil. Mag.*, **40**, 394.
 LEPRINCE-RINGUET and HEIDMANN; LI and PERKINS, 1948, *Nature, Lond.*, **161**, 844.
 PERKINS, D. H., 1947, *Nature, Lond.*, **160**, 707.
 PERKINS, D. H., 1948, *Bristol Symposium*.
 PERKINS, D. H., 1949, *Phil. Mag.* (in publication).
 POWELL, W. M., 1946, *Phys. Rev.*, **69**, 385.
 ROSS, B., 1948, *Rev. Mod. Phys.*, **20**, 564.
 SIMPSON, J. A., 1948, *Phys. Rev.*, **73**, 1389.
 STETTER, G., and WAMBACHER, H., 1939, *Physik Z.*, **40**, 702.
 YUAN, L. C. L., 1948, *Phys. Rev.*, **74**, 504.

LVI. *Mechanism of π -meson Disintegrations.*

By D. H. PERKINS, Ph.D.*

[Received March 7, 1949.]

[Plates III. & IV.]

ABSTRACT.

The present paper describes observations made on the disintegration stars produced by nuclear capture of slow negative mesons, in nuclear plates exposed to cosmic radiation. These measurements are in confirmation of the hypotheses previously put forward (Heidmann and Leprince-Ringuet 1948, Perkins 1948) to account for the conversion of the meson rest-energy into nuclear excitation energy, namely that this mass energy is transferred into kinetic energy of one or two nucleons, which subsequently excite the nucleus in their passage through it.

Previous measurements of the mass of the star-producing mesons (Lattimore 1948, Goldschmidt-Clermont, King, Muirhead, and Ritson 1948) and evidence from grain-counts described below, indicate that they can all, or nearly all, be identified with the negative π -mesons of mass 285 produced in the Berkeley cyclotron.

§ 1. EXPERIMENTAL.

(a) Identification of nuclei.

IN order to obtain detailed information regarding the mechanism of annihilation of a negative meson captured by a nucleus, it has been found essential to separate the meson stars produced in the light nuclei (C, O, N) in the emulsion, from those in the heavy (Ag, Br). This has been done by comparing the meson stars with "ordinary" stars produced by fast cosmic-ray particles (presumably nucleons), on the assumption that the process of disintegration of a nucleus is independent of the *mechanism* of excitation, *i. e.* that the energy distribution of emitted particles is determined by a single parameter, the nuclear temperature (apart from nuclear constants like mass number and potential barrier).

The total number of meson-produced stars examined to date is 120. The number of stars with different numbers of tracks or prongs (apart from recoil tracks) are given in the second row of the table.

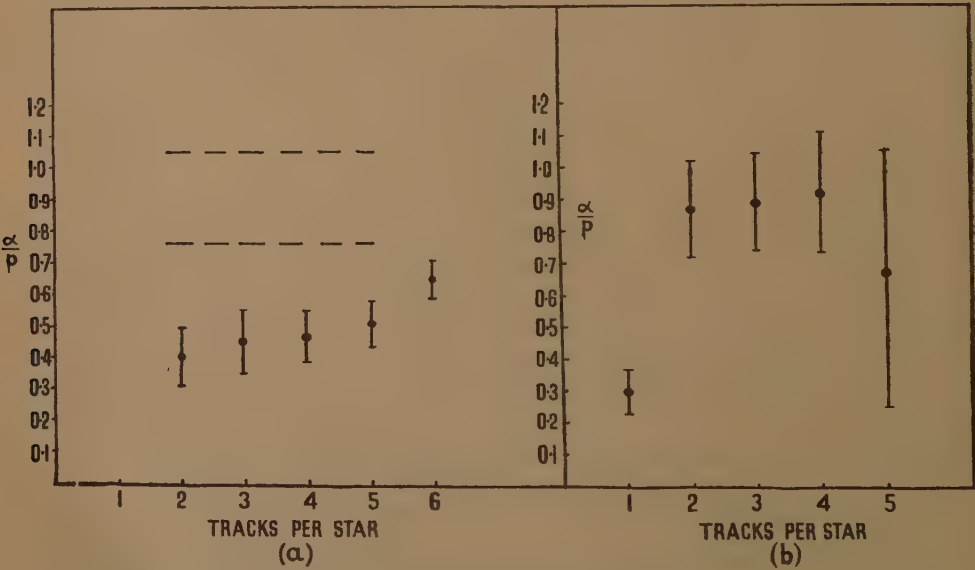
* Communicated by Sir G. Thomson, M.A., D.Sc., F.R.S.

Fig. 1 indicates the ratios (α/p) of α -particles to protons from stars. In actual fact the term "alpha-particles" refers to all tracks, longer than 5μ , produced by particles of charge equal to or greater than two units, and

TABLE.

No. of prongs per star	0 (recoil tracks only)	1	2	3	4	5
No. of stars observed	8	48	30	22	11	1
Proportion of stars showing recoils	All	0.40	0.53	0.23	0	0
α/p ratio	—	0.30 ± 0.07	0.87 ± 0.15	0.89 ± 0.15	0.92 ± 0.19	0.67 ± 0.4
Estimated No. of stars from Ag, Br	30 (see text)	44	5	2	1	0
Estimated No. of stars from C, O, N	0	4	25	20	10	1
True percentage of stars of different prong number (in- cluding hydrogen)	29	30	19	$14\frac{1}{2}$	7	0.6
Berkeley results [†] (Bull. <i>Amer. Phys. Soc.</i> , 1949)	27	23.4	24.0	14.8	8.7	1.9

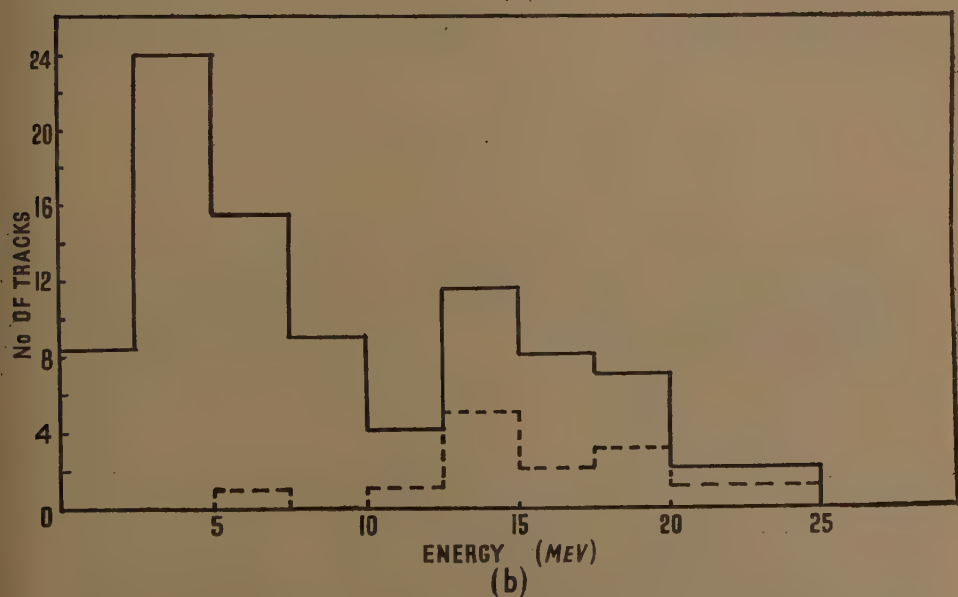
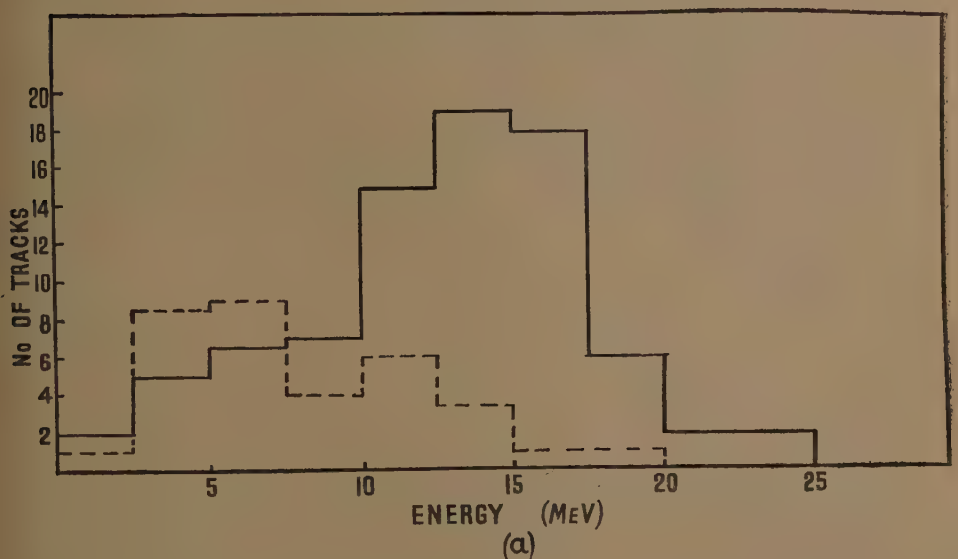
Fig. 1.



- (a) $\frac{1}{2}$ α/p ratio in "ordinary" stars, standard emulsion.
- average α/p ratio from "ordinary" stars in gelatine sandwich.
- (b) α/p ratio in meson stars.

"protons" refers to all heavy singly-charged particles (*i. e.* includes deuterons and tritons as well as protons). The points in fig. 1(a) indicate the α/p ratio for different prong number in "ordinary" stars produced in

Fig. 2.

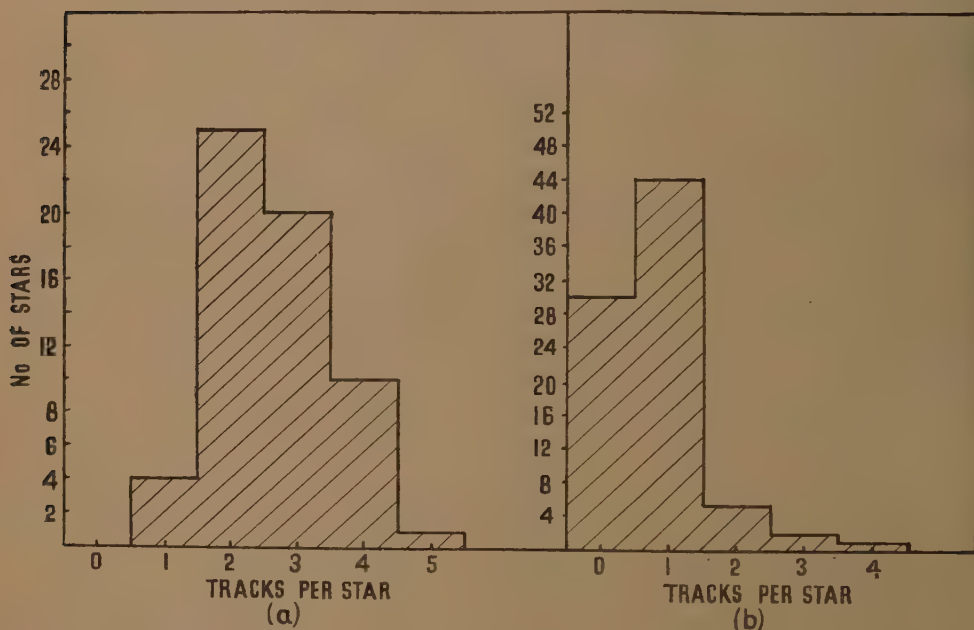


(a) — energy distribution of alpha-particles from "ordinary" stars of 2, 3 or 4 prongs, standard emulsion.
 energy distribution of alpha-particles from "ordinary" stars in gelatine sandwich.

(b) Energy distribution of alpha-particles from meson-stars.
 — all stars. one prong stars.

normal emulsion, 75 per cent of which are due to silver and bromine nuclei. The dotted lines indicate the limits of statistical error for the same ratio obtained from a small number (27) of stars produced in the light nuclei (C, O, N) of thin gelatine layers incorporated in the emulsion. From these data (extrapolating the ratio for stars of two or more prongs back to one-prong) we conclude that, for stars of <5 tracks, the α/p ratio in C, O, or N is ~ 1 , whereas that in silver or bromine is ~ 0.3 —a result we expect from considerations of the relative heights of the potential barriers. The α/p ratios for meson-stars (fig. 1(b)) show therefore that nearly all the one-prong stars are due to heavy nuclei, whereas most of those of two or more prongs are from light nuclei.

Fig. 3.



Size distribution of meson stars in (a) light nuclei (C, O, N) and (b) heavy nuclei (Ag, Br).

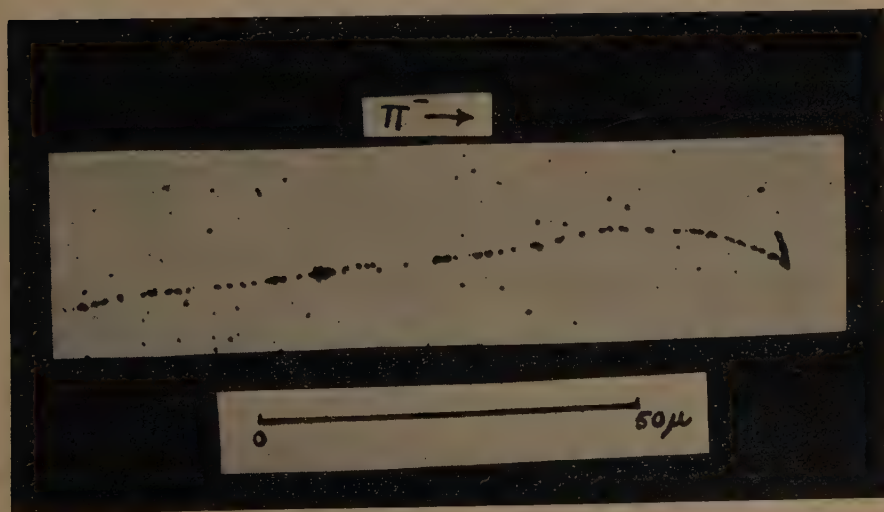
This view is confirmed by measurement of the energies of the emitted alpha-particles. The full-line histogram of fig. 2(a) indicates the energy distribution of alpha-particles from 100 "ordinary" stars of two, three and four prongs (in standard emulsion), and the broken line that from the corresponding number of stars (25) in pure gelatine. The difference of the two distributions will be the energy spectrum of alpha-particles from the heavy nuclei (Ag and Br) only. It is evident that practically all alpha-particles of energies below 10 MeV. can be attributed to light nuclei. The distribution of alpha-particles from meson-stars (fig. 2(b)) shows that most of the alpha-particles originate from light nuclei; there is some evidence

FIG. 4.



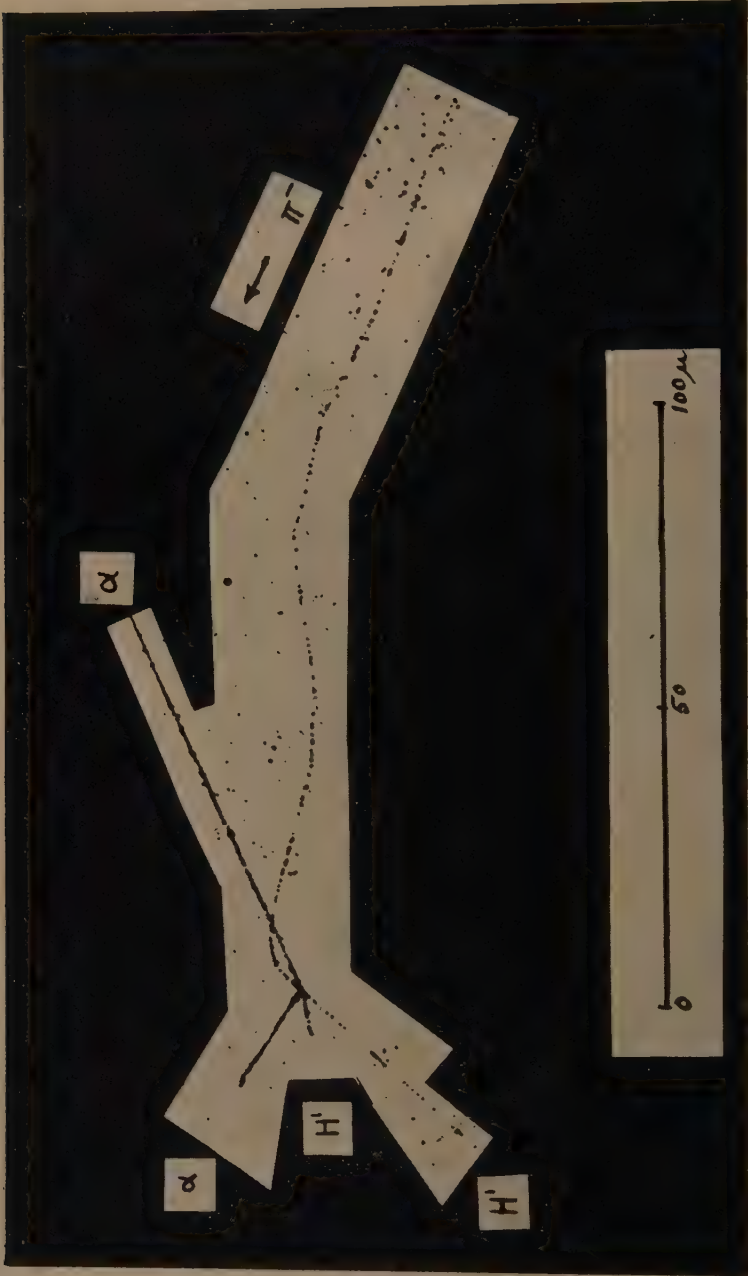
Complete disintegration of carbon nucleus by meson capture. One of the alpha-particle tracks appears very thick owing to its steep angle of dip in the emulsion. Ilford C2 plate, exposed at Jungfraujoch (3500 m.).

FIG. 7.



Example of π -meson producing a short nuclear recoil track on coming to rest in the emulsion. Ilford C2 plate. Exposure at La Oroya, Peru (5500 m.).

FIG. 9.

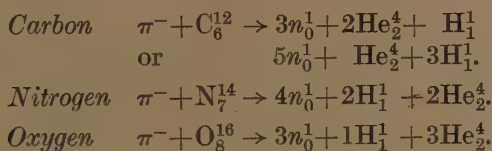


Disintegration of nitrogen nucleus by π -meson. Ilford C2 plate, exposed at sea-level (London).

for a small second peak due to those alphas from Ag and Br nuclei. The dashed histogram of fig. 2(b) corresponds to energies of alpha-particles from one-prong meson stars only, again showing that most of these are from heavy nuclei.

From the energies of the α -particles in individual stars, and the average α/p ratios given above, one can distinguish roughly between stars produced in light or heavy nuclei. (See Table and fig. 3.) As an example of the method, we describe the procedure for two-prong stars. From a total of 30 observed, we obtain 28 α -particles and 32 protons. 13 stars contain α -particles of energies below 10 MeV., and hence attributable to light nuclei. These 13 stars have, altogether, 19 α -particles, so that the number of protons definitely from light nuclei will also be about 19 ($\alpha/p \sim 1$). Thus of the total of 32 protons about 13 can be ascribed either to stars in heavy nuclei, or to those in light nuclei which emit α -particles of energies above 10 MeV., or in which the α -particles have undetermined energies owing to the fact that they pass out of the emulsion layer. The number of α -particles in these categories is $(28-19)=9$. The α/p ratio $(9/13)$ for these stars not definitely attributable to light nuclei then leads to a rough estimate of five stars due to Ag or Br.

The first column in the Table was obtained from data described below. The size distribution of meson stars from light nuclei (Table, fig. 3) indicates that in many cases meson capture leads to complete disintegration of the nucleus into alpha-particles, protons and neutrons. Typical disintegrations which have been identified are



Examples of carbon and nitrogen disintegrations are given in figs. 4 (Pl. III.) and 9 (Pl. IV.).

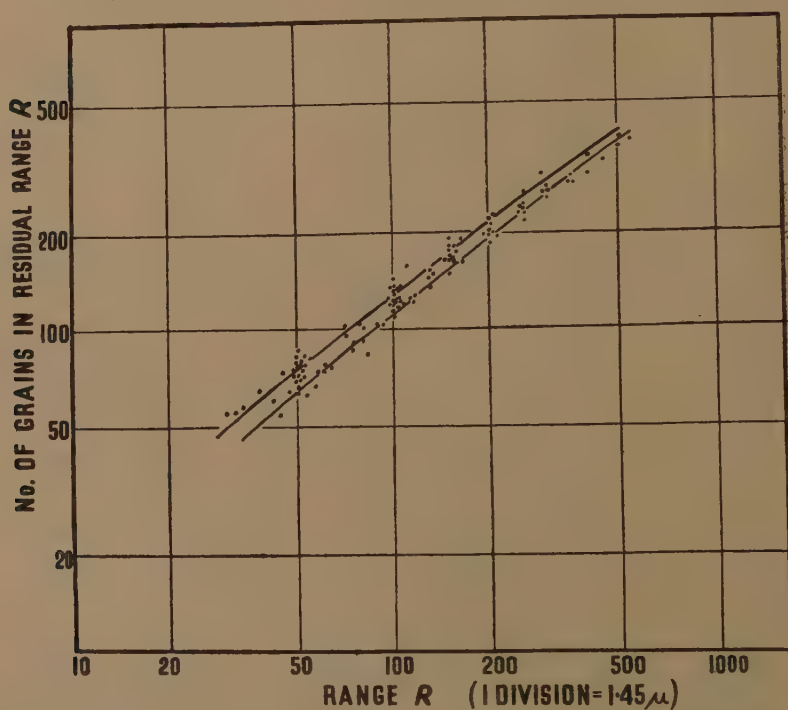
From the Q values of these reactions and the kinetic energies of the ejected particles, we estimate the most probable value of the energy required to liberate the charged particles to be ~ 35 MeV. This value is in good agreement with that obtained by Heidmann and Leprince-Ringuet (1948) and, as has previously been pointed out, is quite small compared with the rest-energy of the π -meson (~ 140 MeV.).

(b) Proton distribution.

The energies of protons ejected in meson stars were determined by grain-counts along the tracks. (A small fraction of the protons come to rest in the emulsion, enabling the energy to be determined by a range measurement.) As has often been emphasized, the results of grain-counting along the tracks in nuclear emulsions are subject to some uncertainty due to latent-image fading. The fading effect was tested by grain-counts along the meson tracks. The great majority of our meson

stars were obtained in plates exposed at low temperature in the ice of a glacier at Jungfrauoch. In fig. 5 have been plotted the results of grain-counting along the tracks of 60 star-producing mesons in these plates. The two curves correspond to the expected limits of probable error ($\pm 0.67\sqrt{n}$) about the mean grain-count n . The ratio of the number of points inside and outside the curves is 1.02 ± 0.14 . This result indicates, first that practically all the mesons can be identified with π -mesons of mass $285m_e$ produced at Berkeley, and secondly that the effect of fading on the grain-count is small compared with the statistical error arising from the finite number of grains along the tracks. Thus we can place reliance on the results of proton energy determination by grain-counting.

Fig. 5.



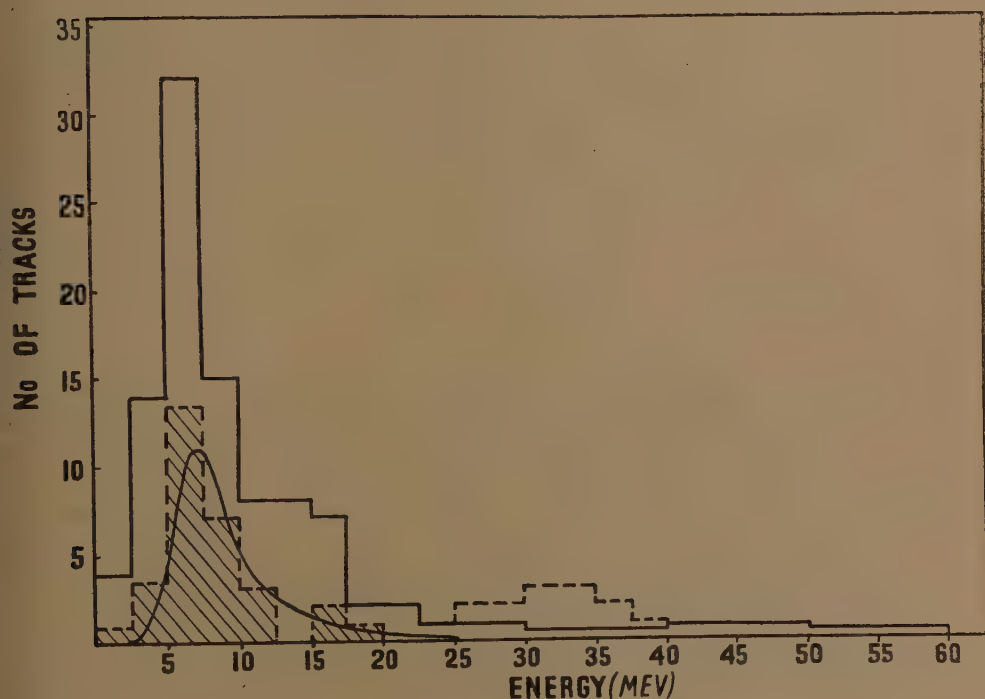
Grain-counts along tracks of star-producing mesons. Ilford C2 emulsion.

The energy distribution of 106 protons from all the meson stars is shown in fig. 6; it is characteristic of a process of nuclear "evaporation", supporting our original assumption that the energy distribution of particles emitted from a given nucleus depends on the excitation energy only, and not on the mechanism of excitation. The broken-line histogram is from the tracks of protons from one-prong meson stars only. It shows two well-separated groups. The shaded one has a peak at about 7 MeV.; the other group (unshaded) corresponds to nine cases of the production by π -mesons of single lightly-ionizing particles, which, if protons, would have energies

lying between 25 and 40 MeV. These are, however, almost certainly examples of π - μ decay in which the path of the μ -meson in the emulsion layer is too short to identify it from the Coulomb scattering, and they have accordingly been omitted from the full-line histogram and the Table.

The full line proton distribution is strikingly similar to the ones obtained from "ordinary" cosmic-ray stars (Perkins 1947), not only in being of roughly Maxwellian form at low energies, but also in exhibiting a high-energy "tail" above 30 MeV., not explicable by an evaporation process. There is strong evidence that the "ordinary" stars are produced by very fast nucleons, which, in addition to "warming up" the nucleus in their passage through it, knock out faster protons (and presumably neutrons) in a Heisenberg process, so producing the "tail" to the spectrum. The

Fig. 6.



Energy distribution of protons from meson-stars. ——— all stars;
 // // // one-prong stars; incomplete π - μ events (see text).
 The curve corresponds to the evaporation distribution from a heavy nucleus of initial temperature $2\frac{1}{2}$ MeV.

energy distribution of knock-on protons should then be determined by the range of force between neutron and proton (Bagge 1944, Rosenfeld 1948). The value obtained from "ordinary" stars is 2.0×10^{-13} cm. (Harding, Lattimore, and Perkins 1949), whilst the proton distribution from meson stars gives 1.7×10^{-13} cm. Further, the average number of fast protons

(above 30 MeV.) per meson star is 0.07 ± 0.02 , as compared with 0.05 ± 0.02 per "ordinary" star at low excitation energies (~ 100 MeV.). Thus, if one assumes that an "ordinary" star is produced by a single fast neutron, we can infer that one, or possibly two, high-energy "primary" neutrons are present in each meson-disintegration. Such neutrons must have energies of at least 60 MeV. (the end-point of the "tail").

(c) *Recoil tracks.*

In addition to the 112 mesons which produce protons and alpha-particles following nuclear capture, we have observed eight examples of mesons which give rise to short nuclear recoil fragments (and also presumably neutrons *) at the ends of their ranges. An example is shown in fig. 7 (Pl. III.).

The statistical error on the mean grain-count along such a small number of tracks does not allow us to attribute them definitely to π -mesons (mass 285) rather than μ -mesons (mass 215). That they almost certainly belong to the former class is shown by the following evidence. A comparison has been made of the number of single (ρ), star-producing (π^-) and recoil-producing mesons in plates at the surface of, and at a considerable depth in, the ice of a glacier at Jungfrauoch :—

	Near surface	4-6 metres
ρ	136	141
π^-	38	10
Recoils	4	0

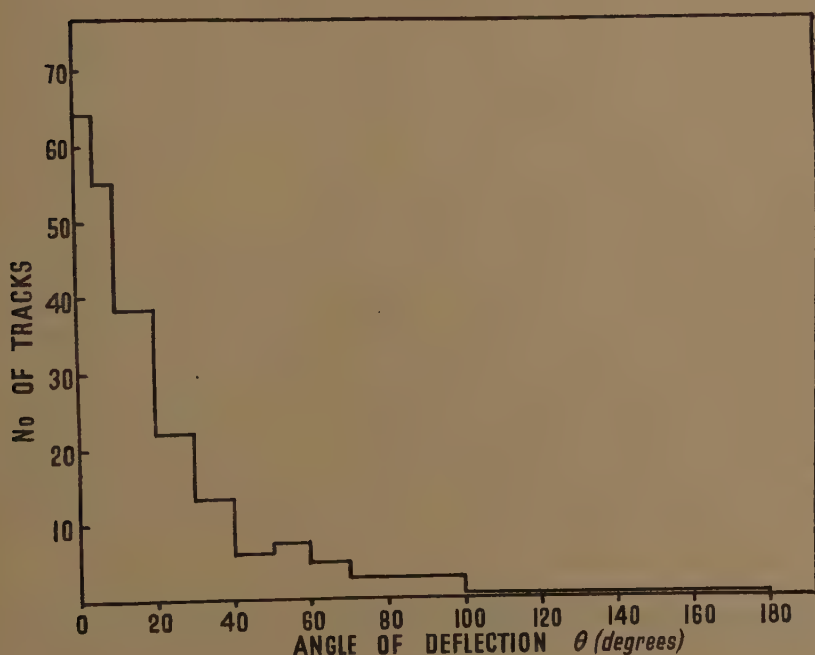
Thus, as far as the poor statistics allow, one can say that the number of mesons showing recoil tracks is proportional to the number of π^- mesons present, and not the ρ -mesons (known to be predominantly μ -mesons of mass 215 (Goldschmidt-Clermont *et al.* 1948)). It can be seen from the size distribution of fig. 3 that the number of cases in which π^- -mesons undergo capture in C, O, or N and lead to emission of neutrons only must be very small, and we assume as a first approximation that the single recoil fragments are from heavy nuclei only. We have estimated their true number by two methods.

(i) From the Table we see that the efficiency for observing recoil tracks from the one-prong stars (practically all heavy nuclei) is only ~ 0.4 ; the remainder are too short to detect. If we assume (rather crudely) that the probability of detecting a single recoil track at the end of a meson track is equal to that for observing a recoil in a star, then the estimated number of π^- -mesons giving recoils only will be about 20. This has to be multiplied by a further factor of 1.5 to account for the relative probabilities of observing single and star-producing mesons during searching of the plates. We thus obtain a total of $30\pi^-$ -mesons undergoing capture in Ag and Br and producing fast neutrons only.

* Apart from the "primary" neutrons postulated above.

(ii) This figure has been checked in the following manner. Direct observation of a single recoil is rather difficult, since it can easily be confused with the Coulomb scattering at the very end of the meson track. Measurements have therefore been made on the change of direction θ , in the plane of the emulsion, along the last two or three microns of track length of a sample of 170 single (ρ) mesons. If the deflection is due to Coulomb scattering, the distribution of θ should be approximately Gaussian. If what appears to be the last one or two microns of the track is really a recoil fragment, the angular distribution should, on the contrary, be isotropic. In the histogram of fig. 8, it is found that the isotropic "back-

Fig. 8.



Distribution of Coulomb scattering angle at ends of single meson tracks.

ground" constitutes less than 5 per cent of the total. Hence, from a total of 800 single mesons in the plates, we obtain 40 as the upper limit to the number of π -mesons producing neutrons and recoil fragments only. This figure is in fair agreement with the one above (30), which represents $(30/112 + 30) \simeq 20$ per cent of all the negative π -mesons coming to rest in the emulsion. Considering the crudeness of the assumptions made, this is in good agreement with the Berkeley results (27 per cent) for the proportion of such mesons not giving rise to stars at the ends of their tracks, and it is the value given in the fifth row of the Table and in fig. 3.

§ 2. INTERPRETATION OF RESULTS.

(a) *Relative numbers of stars in light and heavy nuclei.*

A meson stopping in a nuclear emulsion will in general spend the last fraction of a micron of its range either entirely in the neighbourhood of heavy atoms (Ag and Br) or of light ones (C, O, N, H). Thus the probability of the meson stopping in heavy or light atoms will be almost independent of the complicated processes of energy loss below a velocity of about 5×10^8 cm. sec⁻¹ and can be estimated from the stopping power formulæ given by Bethe (1937). The calculated proportion of mesons stopping in light atoms is 45 per cent; the observed fraction of stars produced by mesons in light atoms is 42 per cent. This confirms the Berkeley result, namely that negative μ -mesons (mass 215) do not produce stars when stopped in heavy atoms (as is well known, they undergo spontaneous decay when stopped in light atoms). The negative μ -mesons stopping in heavy atoms in our plates are over twice as plentiful as all the negative π -mesons, and if they gave rise to stars, would seriously reduce the observed fraction of meson stars from light nuclei. We cannot however exclude the possibility that a very small number of star-producing mesons are heavier than π -particles.

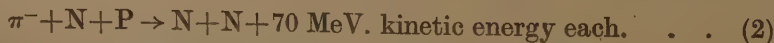
In this analysis, we have not so far taken any account of π^- -mesons which may be captured into the Bohr orbits of hydrogen nuclei in the emulsion. These mesons might, for instance, produce a fast neutron and another neutral particle. From stopping power calculations, and the known emulsion constitution, we estimate that they would constitute about 10 per cent of all the π^- -mesons and would thus increase the proportion of π^- -particles not giving stars to 29 per cent. The proportions of meson stars of different prong number, allowing for the effect of hydrogen, are given in the seventh row of the Table. They appear to agree very well with the results in the Berkeley experiments (*Bull. Amer. Phys. Soc.*, 1948).

(b) *Mechanism of excitation process.*

Examination of the energy distribution of protons from meson stars, described above, suggests that at least one fast neutron (energy above 60 MeV.) results from nuclear capture of a π -meson. It has been suggested that the π -meson, in undergoing annihilation, excites to high energy a single nucleon (Heidmann and Leprince-Ringuet 1948, Perkins 1948)

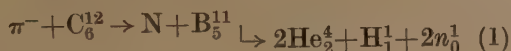


though from the standpoint of conservation of momentum, a more probable process seems to be production of two fast neutrons moving in opposite directions

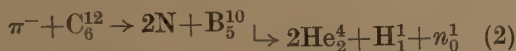


In what follows, we shall term the fast neutrons in (1) and (2) "primary" neutrons (N) to distinguish them from those released in the process of

evaporation (n_0^1). The above reactions in carbon now become



and



where, assuming the evaporated neutrons carry off about the same energy as the protons, the most probable excitation energy produced by the primary neutrons is of the order of 50 MeV. Allowing for the fact that the disintegration schemes given above are from stars of 3, 4 and 5 prongs which can be identified, whereas some of the disintegrations produced in the light nuclei give rise to emission of only one or two charged particles, we arrive at an estimate of 40 MeV. for the average excitation energy produced in a carbon, oxygen or nitrogen nucleus following π -meson capture. This is much less than the total kinetic energy of the primary neutrons, which is what we would expect, for, as pointed out by Serber (1947), neutrons of such high energies in general lose only a small fraction of their energy in traversing a nucleus. The energy loss should then be proportional to the nuclear radius, *i. e.* $A^{\frac{1}{3}}$. Hence the excitation in a silver or bromine nucleus capturing a π^- -meson will be about 80 MeV. Observations on "ordinary" stars in Ag and Br nuclei at different excitation energies U give us a relation between U and the initial temperature T_0 of the nucleus (Harding, Lattimore, and Perkins 1949). For $U=80$ MeV. we estimate $T_0 \sim 2\frac{1}{2}$ MeV. This value is in good agreement with the observed Maxwell energy distribution of protons in one-prong meson stars, and therefore from heavy nuclei (see fig. 6, shaded histogram, and calculated evaporation curve).

The mean free path between collisions in nuclear matter is of the same order of magnitude as the nuclear radius for neutrons of energies ~ 100 MeV. Thus for a given path-length of the neutron through the nucleus, the number of collisions (and also the energy loss per collision) undergo large fluctuations. Averaged over a large number of events however, one can say that corresponding to each path-length there is a certain mean energy loss, proportional to the path-length itself.

The mean free path of energetic neutrons traversing nuclear matter is well known from the experiments at Berkeley. It can be expressed roughly as $\lambda = 4 \times 10^{-15} E$ cm. where E is the energy in MeV. (Serber 1947). Let us assume that, when the π -meson is annihilated, the resulting "primary" neutrons originate at random inside the nuclear volume, and also have random direction. If l is the path-length, assumed rectilinear, r the nuclear radius, and $x = l/2r$, the probability distributions in path-length through the nucleus are

$$w(x) dx = \frac{3}{2} (1-x^2) dx \quad 1 \text{ neutron} \quad (1)$$

$$w(x) dx = 2x dx \quad 2 \text{ neutrons} \quad (2)$$

and the average values are $\bar{x}_1 = 0.375$ and $\bar{x}_2 = 0.67$ respectively. For silver or bromine, $r \sim 1.5 A^{\frac{1}{3}} 10^{-13}$ cm. $= 7 \times 10^{-13}$ cm. Hence $\bar{l}_1 = 5 \times 10^{-13}$

cm., $\bar{l}_2 = 9 \times 10^{-13}$ cm. The energies of the neutrons in (1) and (2) are 140 MeV. and 70 MeV. respectively. Thus $\lambda_1 \sim 6 \times 10^{-13}$ cm., $\lambda_2 \sim 3 \times 10^{-13}$ cm., and, if n_1 and n_2 denote the average number of collisions of the primary neutrons in traversing the nuclei, then $n_1 \sim 1$, $n_2 \sim 3$. Assuming, further (Serber 1947) an average energy transfer of 25 MeV. per collision, we obtain values $\bar{U}_1 \sim 25$ MeV. and $\bar{U}_2 \sim 80$ MeV. for the excitation energies in each case. This calculation is very crude, but the agreement between U_2 and the observed mean excitation energy in AgBr strongly favours the 2-neutron process. It should be pointed out that if, as seems probable, the π -meson is more likely to be annihilated in a volume element near the surface of the nucleus rather than the centre, the value of U_1 will be still further decreased, whilst U_2 will remain the same.

The distribution in path lengths $w(x)$ allows one to calculate, very roughly, the relative number of disintegrations in heavy nuclei leading to evaporation of neutrons only, and those leading to emission of both neutrons and protons. We assume (Harding, Lattimore, and Perkins 1949) that the (Maxwellian) energy distribution of protons evaporating from a heavy nucleus at known temperature can be derived from that of the neutrons by simply multiplying by the barrier penetrability; this gives the ratio of protons to neutrons emitted at a fixed *instantaneous* nuclear temperature, and hence, allowing for "cooling", at a given initial temperature (T_0). The distribution of path-lengths, and therefore initial temperatures, then leads to an estimate of the ratio of the numbers of disintegrations giving charged-particle "stars" and those in which only neutrons are evaporated off. The value obtained for this ratio is about unity, in accordance with the above observations. The calculation cannot, however, be very accurate, since it depends very critically on the choice of the average initial temperature T_0 ($\sim 2\frac{1}{2}$ MeV.), and no allowance has been made for the cut-off in energy transfers below the Fermi energy, or for the fact that the paths of the "primary" neutrons through the nucleus are not rectilinear. Further, no account has been taken of the alpha-particles which are sometimes emitted. Substantially the same result has been obtained using the distribution of excitation energies calculated much more rigorously by Goldberger (1948), for the bombardment of heavy nuclei by 100 MeV. neutrons.

§ 3. CONCLUSIONS.

Our conclusions may be summarized as follows :—

1. Separation of the meson stars into those from heavy and light nuclei confirm the Berkeley result, namely that none of the stars are produced by negative μ -mesons.
2. The energy dissipated in the meson-produced disintegrations is much less than the rest-energy of the meson ($m_\pi \sim 140$ MeV.).
3. Examination of the energy distribution of protons from meson stars indicates that one or two fast neutrons, of energies at least 60 MeV. are probably produced in a primary process by annihilation of the meson.

These lose only a fraction of their energies in traversing the nucleus and producing excitation followed by evaporation. The average energy loss is about what one would expect for neutrons of energy ~ 70 MeV., suggesting that the meson interacts with a neutron-proton pair, giving two fast neutrons moving off in opposite directions.

4. Most of the multi-prong stars due to π^- -mesons arise from light nuclei. In about half the disintegrations following capture of a π^- -meson by a heavy nucleus, a single proton or alpha-particle is emitted, and most of the remainder lead to ejection of neutrons only. Thus about one quarter of all the π -mesons do not produce a visible star on coming to rest in the emulsion.

5. Neutrons with energies ~ 70 MeV. are above the fission threshold in lead, bismuth, etc. Thus meson-induced fission should be a probable process in such nuclei. Examination of plates in which layers of lead phosphate are sandwiched between layers of normal emulsion, is in progress. Such processes will be difficult to observe owing to the small amount of lead which can be incorporated by this method.

ACKNOWLEDGMENTS.

I should like to express my indebtedness to Professor Sir George Thomson for valuable discussions and advice, and to Herr H. Wiederkehr, Dr. R. Stämpfli and Professor A. von Muralt, for providing facilities at the Forschungs station, Jungfrauoch.

Note added in proof.—Recently, more accurate measurements have been made in this laboratory of the thicknesses of the layers in gelatin sandwich plates (see Harding J. B., 1949), *Nature, Lond.*, **163**, 440. The proportion of "ordinary" stars in normal emulsion due to light nuclei has been increased from 25 per cent to 36 per cent. This correction will not effect the main argument in § 1.)

REFERENCES.

- BAGGE, E., 1944, *Phys. Z.*, **21**, 461.
 BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9**, 70.
Bulletin American Phys. Soc., 1949, Feb. 3rd.
 GOLDBERGER, M., 1948, *Phys. Rev.*, **74**, 1269.
 GOLDSCHMIDT-CLERMONT, KING, MUIRHEAD, and RITSON, 1948, *Proc. Phys. Soc.*, **61**, 183.
 HARDING, J. LATTIMORE, and PERKINS, 1949, *Proc. Roy. Soc.* (in the press).
 HEIDMANN, J., and LEPRINCE-RINGUET, L., 1948, *C.R. Acad. Sci., Paris*, **226**, 1716.
 LATTIMORE, S., 1948, *Nature, Lond.*, **161**, 518.
 PERKINS, D. H., 1947, *Nature, Lond.*, **160**, 299; 1948, *Ibid.*, **161**, 487.
 ROSENFELD, L., 1948, *Nuclear Forces*, Vol. I (Amsterdam: North Holland Publishing Co.).
 SERBER, R., 1947, *Phys. Rev.*, **72**, 1114.

LVII. *The Diffraction of Radio Waves from Meteor Trails and the Measurement of Meteor Velocities.*

By J. G. DAVIES and C. D. ELLYETT,
Physical Laboratory, University of Manchester*.

[Received March 15, 1949.]

[Plate V.]

ABSTRACT.

The necessity for studying the reflected amplitude of radio waves from meteor trails in time intervals of the order of milliseconds is discussed, and automatic equipment is described, using pulse technique, which records this information photographically. Theoretical considerations indicate that fluctuations in reflected amplitude should occur as the meteor crosses the region of the perpendicular from the observing station, the problem being analytically identical to the diffraction of light at a straight edge. It is shown experimentally that initial amplitude fluctuations are associated with many meteor echoes and possess the correct ratios of zone durations predicted by the diffraction theory. Since the times to traverse specified zones are measured, the experiment leads directly to a measure of the velocity of individual meteors. The technique has been applied to the velocity measurement of the meteors from known showers and the results are shown to agree with earlier photographic and visual measurements.

INTRODUCTION.

IN the original work on the radio detection of meteor trails (Hey and Stewart 1946; Prentice, Lovell and Banwell 1947; Lovell, Banwell and Clegg 1947) the range, amplitude and duration of the radio echoes were obtained from visual or photographic observations of conventional cathode ray tube displays. Such methods give little information about the processes occurring during the actual formation of the column of ionization as the meteor sweeps into the earth's atmosphere. Moreover, the early observations indicated that the radio echo from the trail was very complex in character (Lovell, Banwell and Clegg 1947), and that a method of studying the formation of the column in detail, and its subsequent history during intervals of the order of one millisecond, would be essential to a fuller understanding of the behaviour of meteors.

* Communicated by Dr. A. C. B. Lovell.

Further, Herlofson (1948) suggested that during the process of formation the radio waves should be scattered from the lengthening column of ionization, giving a diffraction pattern analagous to the diffraction of light at a straight edge. Successful observation of such a pattern would make it possible to calculate the velocity of the meteor from the zone spacings.

Part I. of this paper describes the apparatus which has been developed to study the amplitude of the reflected radio wave at intervals of 1.6 millisecond. Part II. deals with the application of this technique to the study of the diffraction phenomenon, and in Part III. the method is applied to the measurement of meteor velocities.

The technique also gives detailed information about the complex amplitude variations which occur after the primary process of formation of the ionized column is completed, and provides a method for studying short-lived radio echoes, with durations of the reflected wave down to 0.005 seconds. These aspects of the subject will be described in later publications.

PART I.

AN AUTOMATIC RECORDER FOR TRANSIENT RADIO ECHOES.

1. *Design Requirements.*

In order to obtain detailed information of the character of transient echoes from meteor trails it is necessary to record automatically the range, amplitude, duration and variations in strength of the echo during the first 0.1 seconds of its life. Among the variations expected are those due to diffraction, discussed in Part II., requiring measurements of received amplitude at millisecond intervals. Signal strength variations in such short time intervals can be satisfactorily studied by recording the amplitude of successive received pulses, using pulse recurrence frequencies of the order of 500 to 1000 per second.

The use of such high pulse recurrence frequencies introduces a difficulty in range measurement since meteor echoes are commonly observed at ranges up to 900 km. Thus the time interval between transmitting a pulse and receiving the echo may be several times the interval between successive pulses. Under these conditions it is not possible to determine to which transmitter pulse a given echo is related. This ambiguity has been overcome by transmitting every fourth pulse as a pair, separated by 200μ sec. Every fourth echo thus appears as a double, which enables it to be associated with the previous double transmitter pulse and so establishes the range without ambiguity.

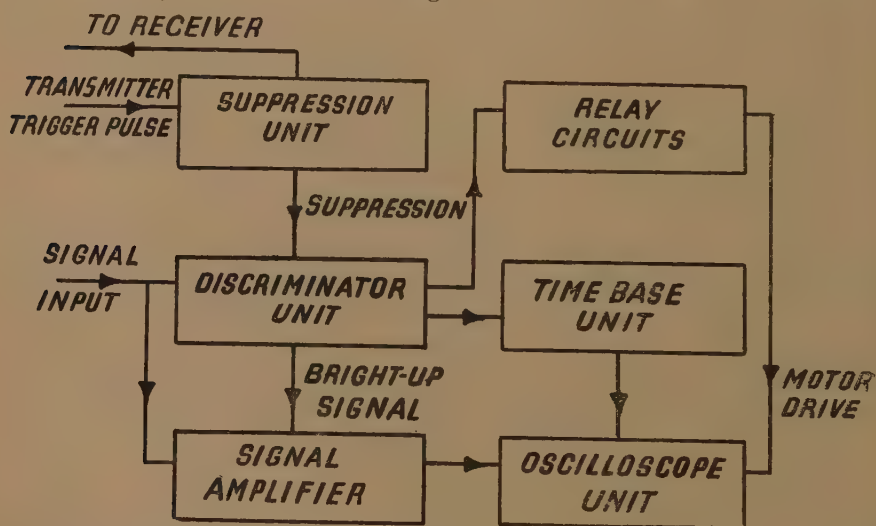
Since measurements are required within the first few milliseconds of the life of an echo, the recording apparatus must be triggered by the echo itself, with a minimum delay. This introduces the problem of preventing the apparatus from being triggered by random noise impulses caused by atmospherics, ignition systems, etc., while keeping it sensitive to echoes. The duration of an echo pulse is determined by the duration of the

transmitter pulse, while that of a noise impulse is inversely proportional to the bandwidth of the receiver. In standard radar techniques the receiver bandwidth is reduced until the duration of a noise impulse is equal to that of the transmitter pulse, this giving the condition of maximum sensitivity. In order to distinguish between noise impulses and echoes, however, this condition has been sacrificed and the receiver bandwidth increased about three times, so that the duration of noise impulses is about 3μ secs, while the transmitter pulse length, and hence echo length, is 8μ secs. This fact is used to discriminate between noise and echoes, by a method described in detail in 3 below.

2. General Description of Apparatus.

The recording unit consists essentially of two cathode ray tubes, with single stroke time bases initiated by the first echo pulse received. The

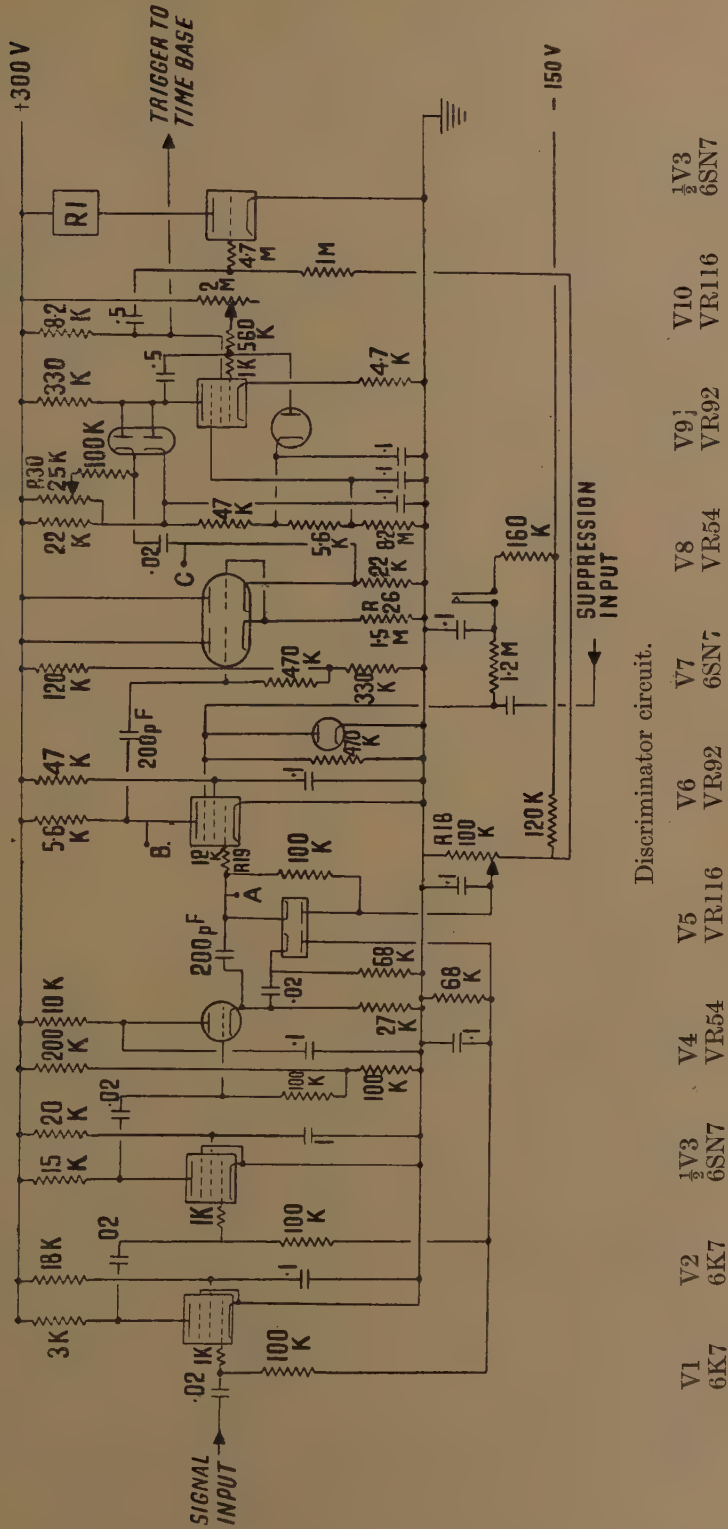
Fig. 1.



first time base, of 7 milliseconds duration, displays the four echo pulses following the one that initiated the time base, together with the associated transmitter pulses. From this, the range of the echo can be measured. The second time base lasts for 0.1 second, thus displaying the first 60 echo pulses; or if the echo lasts for less than 0.1 second, the full history of the echo.

The output of the receiver is fed into the discriminator unit (fig. 1), which separates echoes from noise impulses, and provides a triggering waveform for the time bases and relay circuits. The relay circuits operate the camera motor immediately after each photograph is taken, and render the apparatus insensitive to further triggering impulses until the next frame is in position. This operation takes 1.5 seconds, and if an echo lasts for more than this time, further photographs will be taken at 1.5 second intervals until the end of the echo.

Fig. 2.

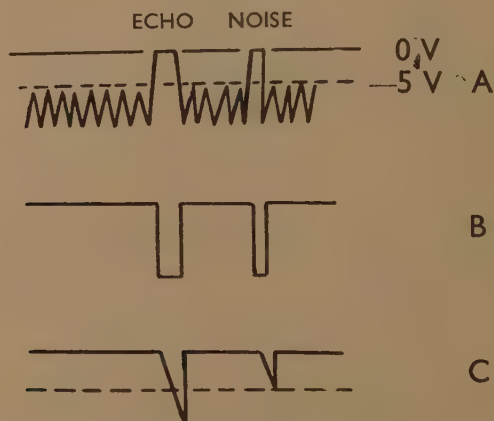


The suppression unit provides pulses which make the receiver and discriminator insensitive for the period of the transmitter pulse and a short interval afterwards, thus preventing echoes from nearby ground sources triggering the apparatus. With the exception of the discriminator the circuits are of standard design and will not be described in detail.

3. Discriminator Circuit.

The circuit diagram is given in fig. 2. Valves V1, V2 and half of V4 form a video amplifier, incorporating automatic volume control. Since the load presented by V5 is rather heavy on certain parts of the waveform, it has been necessary to introduce the cathode follower V3. The output at the cathode of V3, due to the normal noise level of the receiver, is about 15 volts peak to peak. The signal at the input is so phased that echoes are positive, hence signals on the cathode of V3 are positive also. The bias for V5, obtained from a 150 volt negative supply via the potentiometer R18, is normally adjusted to about -20 volts. Hence

Fig. 3.



Discriminator wave forms.

the valve is normally cut off, but begins to conduct as soon as any signal rises much above the receiver noise level. If the signal amplitude reaches about twice noise level, grid current flows in V5, and owing to the large grid stopper R19, the amplitude is effectively limited at this value. Thus on the anode of V5 (*B* of fig. 3) the receiver noise has been eliminated from the signal, and all signals have been reduced to a constant amplitude.

This signal is fed to the grid of the first half of V7, which is returned to a potential of about $+220$ volts. The cathode resistor R26 is 1.5 M , and hence very little current flows in the stage. The cathode therefore rises to 230 volts, the valve current being about $150\mu\text{ A}$. As soon as a signal is received on the grid, the valve becomes completely cut off, and the cathode starts to fall towards zero volts, at a rate determined by R26 and the stray capacities on the cathode. On the second edge of the pulse, however, the grid rises rapidly to its original potential, and the valve draws a large current, rapidly recharging the cathode to the original

voltage. Since the effective impedance at the cathode over nearly all this rise is of the order of 1000 ohms, the rate of the rise is about 1500 times greater than the downward slope of the cathode waveform, and is therefore practically instantaneous. The stage is then ready to receive another pulse immediately. This is of importance, as interference caused by ignition systems frequently consists of a train of impulses with very little interval between them.

During an echo the cathode will have had time to fall through about three times the voltage which it falls through during a noise impulse, since the transmitter pulse length is about three times the length of a noise impulse. With the valves used these are about 30 volts and 10 volts respectively. The cathode is directly coupled to the grid of the second half of V7, which is a conventional cathode follower, required because of the very high output impedance of the previous stage. The effect of the circuit to this point has been first to limit the amplitude of all signals to a constant value, and then to make the amplitude of the signals proportional to their duration.

The output of V7 is used to trigger a phantastron circuit V10 via the diodes V8 and potentiometer R30, the latter controlling the signal amplitude required to trigger the phantastron. This potentiometer is adjusted so that the circuit is triggered by a signal of about 8μ sec. duration, thus discriminating between noise impulses and echoes. The delay time of the phantastron is of the order of 0.2 seconds. The positive square wave from the screen grid is used to trigger the time base circuits, and is also applied to the grid of a triode ($\frac{1}{2}$ V3) in the anode circuit of which is the relay coil R1.

4. Operational Results.

The efficiency of the above circuits in discriminating against unwanted noise is illustrated by the following examples of echoes recorded under two very different sets of conditions. First, meteors of known showers, with a high rate—20 to 100 echoes per hour—at night, under quiet noise conditions; and secondly, non-shower, or sporadic meteors—where the rate is very low, rarely over 10 per hour—under conditions of bad noise interference.

During the Geminid shower of 1948, December 11–14, the apparatus was operated for a total of 27 hours 12 minutes, in which time 2585 frames were taken. Fifty-five per cent of these frames were triggered by echoes, while 45 per cent were triggered spuriously.

In 243 hours 36 minutes observation on random meteors between 1948, September 18, and 1948, December 22; 1795 frames were triggered by echoes out of a total of 14,536 frames. During much of this time intense interference was encountered from lorries in the immediate vicinity of the apparatus. Thus under bad daytime conditions, with a mean echo rate of 6.6 per hour, 12.3 per cent of frames were triggered by echoes; while under good night-time conditions, with a mean hourly rate of 44.7, the yield rose to 55 per cent.

PART II.

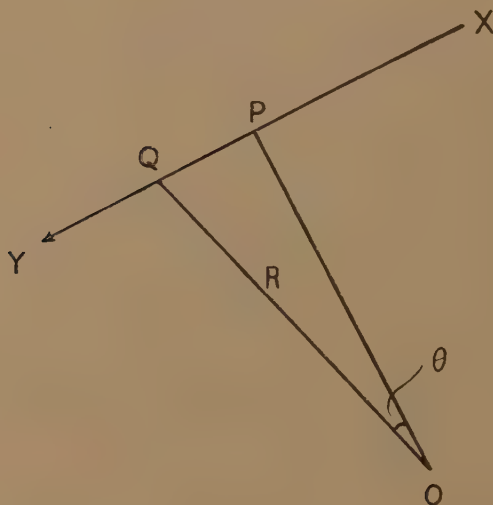
THE DIFFRACTION EFFECT.

1. *Intensity of the Reflected Radio Wave.*

It has been shown by Lovell and Clegg (1948) that the equations applicable to the diffraction of radio waves from the narrow column of ionization formed by a meteor are analogous to those obtained on optical theory for diffraction of light at a straight edge.

In fig. 4 if XY is the ionized column and O the observing station, it was shown that the voltage amplitude from a portion XQ of the trail

Fig. 4.



is given to a close degree of approximation by Fresnel's integral

$$V_{XQ} = k \left(\frac{\lambda}{R} \right)^{3/2} \left[\int_{-\infty}^{\nabla} \cos \frac{1}{2} \pi \nabla^2 d\nabla + i \int_{-\infty}^{\nabla} \sin \frac{1}{2} \pi \nabla^2 d\nabla \right],$$

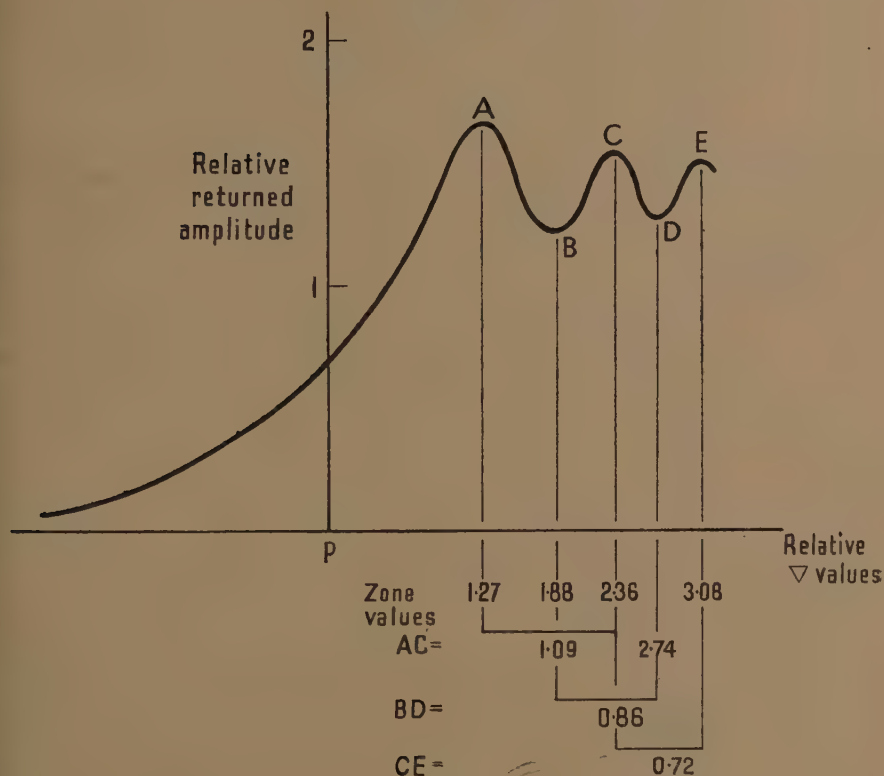
where $\nabla = 2\theta \sqrt{R/\lambda}$. If P is the foot of the perpendicular from the station to the trail, the distance PQ is given by

$$s = R\theta = \frac{1}{2} \nabla \sqrt{R\lambda}. \quad (1)$$

As the meteor sweeps through the atmosphere in the direction XY, the value of ∇ will change from $-\infty$ to $+\infty$ and the echo amplitude should vary in the manner shown in fig. 5, rising as the meteor passes P and showing the characteristic fluctuations as it moves through the first few zones beyond this point. It is clear from the shape of fig. 5 that the echo amplitude should be inappreciable until the meteor is quite close to the perpendicular point P from the observing station. If the abscissa of this diagram be now considered as a time axis, and the meteor velocity is uniform, then the envelope of successive echo pulse amplitudes appearing across the oscilloscope screen (Part I.) should give the time intervals of fig. 5 exactly.

There are a number of reasons, such as non-uniform ionization along the trail of the meteor, which may be expected to invalidate the Fresnel

Fig. 5.



amplitude relationships in many cases. The periods of the successive zones should, however, be determined by the maxima and minima of the curve, independent of the relative amplitudes.

2. Experimental Verification of the Diffraction Formula.

An attempt to verify the above prediction, using the apparatus described in Part I., was made from the photographic records of meteor echoes obtained during the Geminid shower of 1947, December, and subsequent showers in 1948. In Table I. the theoretical ratios of various zone widths are given, together with the equivalent ratios of a number of typical experimentally measured meteors, at various ranges. A, C, E are successive maxima, and B, D, F successive minima, as given in fig. 5. On most echoes there was a small pulse-to-pulse random amplitude variation. Before extracting zone widths, therefore, all curves were first smoothed by averaging with a sliding group of three pulses. In addition, in order to eliminate the possibility of error when the axis of the smoothed oscillatory curve departed from the horizontal, measurements were always made between either adjacent maxima AC CE, etc., or adjacent minima BD, DF, etc.

TABLE I.

Ratios of Zones	AC/DF	AC/CE	BD/DF	AC/BD	BD/CE	CE/DF		
Theo- retical ratios	1.68	1.52	1.32	1.27	1.19	1.11		
Experimental ratios							Range (Km.)	Date 1948
(a) Quadran- tid meteors								
1	1.59	1.56	1.16	1.37	1.14	1.01	360	Jan. 4
2	1.57	1.50	1.38	1.13	1.32	1.05	110	Jan. 4
3	—	1.48	—	1.28	1.15	—	470	Jan. 4
4	—	1.43	—	1.25	1.14	—	360	Jan. 4
5	1.88	1.61	1.36	1.38	1.16	1.17	400	Jan. 4
Mean of Quadran- tids	1.68	1.52	1.30	1.28	1.18	1.08		
(b) Summer daylight meteors								
1	1.69	1.52	1.28	1.32	1.15	1.11	355	June 25
2	1.75	—	1.37	1.27	—	—	590	June 27
3	1.58	1.46	1.18	1.33	1.10	1.08	230	June 28
4	1.66	1.63	1.30	1.27	1.27	1.02	305	June 28
5	1.76	1.66	1.40	1.26	1.32	1.06	375	July 2
6	1.78	1.63	1.27	1.41	1.16	1.09	290	July 2
7	1.53	1.47	1.23	1.24	1.18	1.04	400	July 28
Mean of Summer meteors	1.68	1.56	1.29	1.30	1.20	1.07		
(c) Sporadic meteors								
1	—	1.50	—	1.25	1.20	—	340	Sept. 20
2	—	1.50	—	1.33	1.13	—	320	Sept. 29
3	1.67	1.57	1.42	1.17	1.34	1.06	215	Oct. 28
4	1.71	1.50	1.29	1.33	1.12	1.14	420	Oct. 28
5	1.76	1.50	1.38	1.27	1.18	1.17	220	Oct. 30
6	1.77	1.58	1.32	1.33	1.18	1.12	235	Nov. 3
7	—	1.50	—	1.31	1.15	—	380	Nov. 30
8	—	1.55	—	1.30	1.19	—	365	Dec. 17
Mean of sporadic meteors	1.73	1.53	1.35	1.29	1.19	1.12		



FIG. 6.

1948 June 27^d 09^h 20^m U.T.

Meteor obtained during Summer daylight showers. Range=395 km.

Apparent geocentric velocity= 32.4 ± 3.3 km/sec.

[First four pulses, where the time-base is non-linear, have been brightened for reproduction.]

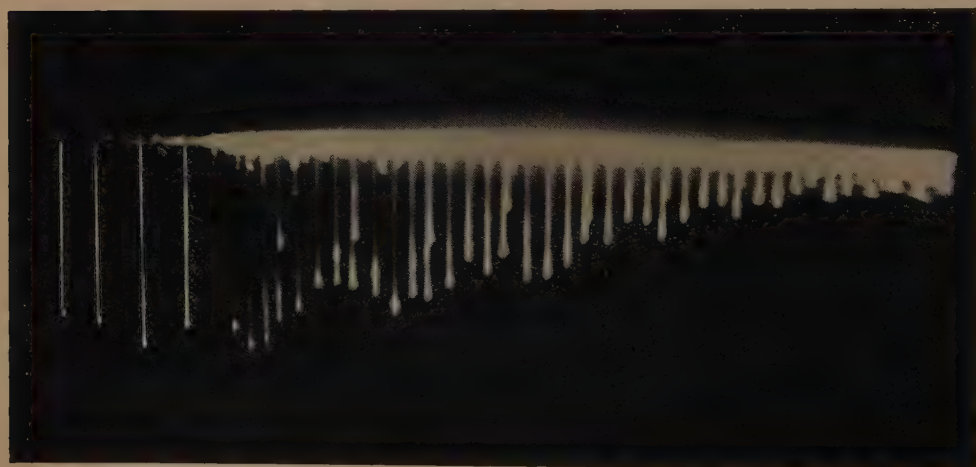


FIG. 7.

1948, June 25^d 11^h 00^m U.T.

Meteor obtained during Summer daylight showers. Range=395 km.

Apparent geocentric velocity= 43.6 ± 2.3 km/sec.

[First four pulses, where the time-base is non-linear, have been brightened for reproduction.]

Within the limits of experimental error, the ratios of the zone spacings in these cases agree with the theoretical values. Two photographs of echoes showing the Fresnel pattern are reproduced in figs. 6 and 7 (Pl. V.).

A second independent criterion can be obtained by studying the phenomenon for a homogenous group of meteors observed at different ranges.

For example,

if V_0 = apparent geocentric velocity of the meteors in a homogenous group.

p = number of pulses observed between specific zone maxima or minima.

t = time between specified zone maxima or minima.

N = pulse recurrence frequency.

Then from equation 1 :—

$$V_0 = \frac{s}{t} = \frac{\nabla}{2} N \frac{\sqrt{R\lambda}}{p} \dots \dots \dots (2)$$

Thus for a series of meteors with identical velocities, $\nabla \sqrt{R}/p = \text{Const.}$ If a number of such meteors are observed at various ranges, widely varying values should be obtained for $(\nabla/p)_{\text{mean}}$ —which is a measure of the reciprocal of the time required to traverse a Fresnel zone—but the values of $(\nabla/p)_{\text{mean}} \cdot \sqrt{R}$ should remain constant.

Some experimental results illustrating these relations are given in Table II. for meteors measured during the Geminid meteor shower of 1947, December 11–14. The value of $(\nabla/p)_{\text{mean}} \sqrt{R}$ is seen to be constant for widely varying values of ∇/p .

TABLE II.

Geminid meteor echoes—1947, December.

Range of echo (km)	$\left(\frac{\nabla}{p}\right)_{\text{mean}} \times 10^{-3}$	$\left(\frac{\nabla}{p}\right)_{\text{mean}} \sqrt{R}$
460	7.85	1.68
170	13.60	1.77
370	8.95	1.72
370	8.84	1.69
420	8.32	1.71
400	8.74	1.75
440	7.90	1.66
220	11.32	1.68
400	8.07	1.61
490	8.30	1.83

This result, and the constancy of the zone ratios in Table I., indicates that the observed diffraction of radio waves from meteor trails obeys optical laws.

3. *The Proportion of Results showing Diffraction Effects.*

An analysis was made of all the echoes automatically recorded during three meteor showers in order to determine the proportion exhibiting Fresnel diffraction curves. The results of this analysis showed that approximately 50 per cent of all echoes exhibited a random amplitude fluctuation; which was either of small amplitude, from pulse to pulse; or of large amplitude with a relatively slow period. The probable origin of these fluctuations, which frequently obscured any possible Fresnel zone picture at the beginning of the echo, will be considered in a later paper.

Another 15 per cent of echoes in which zones could not be discerned comprised those cases where either the echo lasted for less than 20 pulses, or only the crests of the oscillations appeared above the noise level.

Approximately 26 per cent of echoes were lost due to instrumental defects, chief of which were uncertain range and range obscured by the transmitter pulse suppression interval. For the first two of the three showers analysed, automatic measurement of range, as discussed in Part I., had not been initiated, the range tube being observed visually. In subsequent showers this loss has been eliminated.

A mean of 6.3 per cent of the echoes gave clear pictures of the Fresnel diffraction pattern. It appears therefore that with optimum conditions in future showers the yield of echoes showing diffraction zones should be of the order of 10 per cent of all echoes recorded.

PART III.

APPLICATION TO THE MEASUREMENT OF METEOR VELOCITIES

The first radio determination of meteor velocities was made by Hey, Parsons and Stewart (1947). Pulsed radio waves at a frequency of 60 Mc/s were transmitted, and the echo, presented by intensity modulation on a cathode ray tube, was photographed on continuously moving film. On 22 occasions a faint arc was discerned before the onset of the main echo. This arc was attributed to scattering from the head of the approaching meteor, the rate of change of range expressing the velocity.

The present diffraction technique was first applied as an alternative radio method of measuring meteor velocities during the Geminid shower of 1947, December 11-14.

The equipment has already been described (Part I., and Prentice, Lovell and Banwell 1947; Lovell, Banwell and Clegg 1947). The aerial beam was maintained at 90° elongation to the Geminid radiant. With this equipment the background echo rate, due to non-shower meteors, was about 1.5 per hour (Lovell and Prentice 1948).

Substituting the constants of the apparatus:—

$\lambda = 4.16 \text{ m.}$, and $N = 620.8/\text{sec.}$, in eqn. (2) gives

$$V_0 = 20.032 \nabla \sqrt{R/p} \text{ km/sec.}$$

where ∇ is the constant for the zone selected (fig. 5), and the range R ,

is in km. From the velocities given by each zone, including multiple zone values such as A, E, etc. (fig. 5), the mean apparent geocentric velocity, \bar{V}_0 , and the standard deviation were evaluated. The results are given in Table III.

With the exception of the fourth echo, all the results of Table III. fall into a close group. In view of the known non-shower rate described above, this particular echo may be due to a stray non-Geminid meteor.

Eliminating this echo, the mean velocity of the group is 34.4 ± 1.45 km./sec. Visual observation of the range tube introduces a mean error of ± 20 km. in range, or ± 0.8 km./sec. in velocity, which is included in the final figure for the mean velocity of the group.

The result is in reasonable agreement with the velocity of 35.95 km./sec. obtained by Whipple (1947) from photographic measurements of visible Geminid meteors.

TABLE III.

Geminid meteor velocities, 1947, December.

Time (U.T.)	\bar{V}_0 (km./sec.)	No. of zone values
11 ^d 01 ^h 05 ^m	33.7 ± 3.3	2
13 03 48	35.6 ± 0.9	4
13 21 27	34.6 ± 1.4	4
14 01 08	28.7 ± 1.6	4
14 03 28	34.1 ± 2.1	4
14 04 43	34.2 ± 3.4	6
14 04 57	35.1 ± 2.5	4
14 06 35	33.2 ± 3.3	6
14 06 56	33.8 ± 2.1	4
14 06 58	32.4 ± 4.1	6
14 21 38	36.9 ± 1.4	3

The same technique was applied during the Quadrantid meteor shower of January 4, 1948, when nineteen velocities were obtained. These results appeared to show the possibilities of three groups (Ellyett and Davies 1948), but the evidence is doubtful because of the small number of results. If the standard deviation of each result is taken into account, there appears to be a main group at 34.5 km./sec., with perhaps a subsidiary group at 39.3 km./sec. There is no reliable data with which these velocities can be compared. Over a period of 27 years various observers have found fourteen visual paths by double observation, from which velocities have been calculated (Fisher 1930). Ten of these results fall into a reasonably compact group, with a mean velocity of 38.4 ± 4.6 km./sec., agreeing well with the radio observations, but the group is not compact or large enough to give any indication of sub-groups.

ACKNOWLEDGMENTS.

The work was carried out at the Jodrell Bank Experimental Station of the University of Manchester. We are indebted to Dr. A. C. B. Lovell, Director of the Station, for his constant interest and advice, and to several

of our colleagues—particularly Dr. J. A. Clegg, V. A. Hughes, F. Moran, I. A. Gatenby and A. Aspinall—for help in making the observations.

One of the authors (C. D. E.) wishes to express his thanks to the council of Canterbury University College, Christchurch, New Zealand, for the granting of three years' leave of absence from teaching duties, and to the Imperial Chemical Industries Limited, for the award of a Research Fellowship. The other author (J. G. D.) wishes to thank the Department of Scientific and Industrial Research for the award of a maintenance grant.

REFERENCES.

- ELLYETT, C. D., and DAVIES, J. G., 1948, *Nature, Lond.*, **161**, 596.
 FISHER, W., 1930, *Harvard College Observatory Circ.* No. 346.
 HERLOFSON, N., 1948, *Reports on Progress in Physics*, **11**, 444.
 HEY, J. S., and STEWART, G. S., 1946, *Nature, Lond.*, **158**, 481.
 HEY, J. S., PARSONS, S. J., and STEWART, G. S., 1947, *Mon. Not. R. Astr. Soc.*, **107**, 176.
 LOVELL, A. C. B., BANWELL, C. J., and CLEGG, J. A., 1947, *Mon. Not. R. Astr. Soc.*, **107**, 164.
 LOVELL, A. C. B., and CLEGG, J. A., 1948, *Proc. Phys. Soc.*, **60**, 491.
 LOVELL, A. C. B., and PRENTICE, J. P. M., 1948, *J. Brit. Astron. Assoc.*, **58**, 140.
 PRENTICE, J. P. M., LOVELL, A. C. B., and BANWELL, C. J., 1947, *Mon. Not. R. Astr. Soc.*, **107**, 155.
 WHIPPLE, F. L., 1947., *Proc. Amer. Phil. Soc.*, **91**, 189.

LVIII. *The Derivation of Statistical Expressions from Gibbs' Canonical Ensemble.*

By F. ANSBACHER and W. EHRENBURG,
 Birkbeck College, University of London*.

[Received March 14, 1949.]

SUMMARY.

The origin of Gibbs' canonical ensemble and the associated partition sum is briefly discussed. It is then shown that for assemblies of particles obeying the Boltzmann, Einstein-Bose or Fermi-Dirac conditions the partition sum possesses some simple properties from which the known expressions for the average distribution in energy of particles are derived elementarily without reference to either Stirling's formula or the method of steepest descent.

§ 1.

SINCE the publication of a fundamental paper by Darwin and Fowler (1922), it has been repeatedly pointed out (*e. g.* Fowler 1936, Schrödinger 1946) that expressions for the number of particles in a given state for

* Communicated by the Authors.

assemblies of independent particles can in a valid manner be derived only by the use of Gibbs' ideas of the canonical or micro-canonical ensemble and not by statements on thermodynamical probabilities. The general method is that of steepest descent, although, particularly for the classical case, simpler methods have been suggested (Tolman 1938). It is the object of this note to propose a derivation of these expressions which, although elementary, meets the criticism which has led to the introduction of the method of steepest descent. The results are obtained by the use of certain simple properties of the partition sums.

§ 2.

It may be as well to review here the ideas leading to the canonical ensemble. A thermodynamic body (not in general a phase) is considered as a kind of giant molecule with STATES i and energies E_i , ($E_i \leq E_{i+1}$). A set of independent thermodynamical variables does not fix the STATE uniquely. In fact a particular body under test may still be in any one of a great number of STATES with, in general, a different probability, and the totality of these STATES with an indication of their probability has been called by Gibbs an Ensemble. The ensemble is canonical if the independent thermodynamical variables are temperature and volume.

If two bodies have the same temperature, then no observable change takes place when they are brought into thermal contact; so that each body is represented by the same ensemble before and after the contact has been established. The combined body, of course, must also be represented by an ensemble. Let i, k, l denote the STATE of the first, second and the combined body, and let $P^x(S)$ denote the probability of finding a body x in a STATE S . Then

$$P'''(l=i, k)=P'(i) \cdot P''(k). \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Now the assumption of elementary disorder underlying all statistics may axiomatically be expressed as follows: within a given ensemble P depends only on the energy of the STATE but not on any other coordinates which determine the STATE. Then from (1)

$$P'''(E_i + E_k) = P'(E_i) \cdot P''(E_k), \quad . \quad . \quad . \quad . \quad . \quad . \quad (1 a)$$

which is in general true only if

$$P'(E_i) = c' \exp (\mu E_i),$$

$$P''(E_k) = c'' \exp (\mu E_k).$$

By expressing the thermodynamical variables E and p as averages over the ensemble, comparing the result in general with the Second Law, and with a perfect gas in particular, one determines μ as $-1/kT$ so that on normalization

$$P(E_i) = \frac{1}{Z} \exp (-E_i/kT), \quad \text{where} \quad Z = \sum_{\text{all STATES}} \exp (-E_i/kT) \quad . \quad (2)$$

is called the partition sum. P defines the canonical ensemble of Gibbs.

All thermodynamical data of interest can be expressed in terms of Z or partial differentials of Z , and the expressions obtained are perfectly general, provided that STATES of the body can be defined.

A particular case arises if the body is an assembly of N independent particles, such as a perfect gas, where each particle may be in a state $^* 1, 2, \dots, l, \dots$ with energies $\epsilon_1, \epsilon_2, \dots, \epsilon_l, \dots$ ($\epsilon_i \leq \epsilon_{i+1}$). Let then n_i^i denote the number of particles in a state with energy ϵ_l when the assembly is in the STATE i . Then

$$E_i = n_1^i \epsilon_1 + n_2^i \epsilon_2 + \dots + n_l^i \epsilon_l + \dots \quad (3)$$

$$N = \sum_l n_l^i \quad \text{for all } i \quad (3a)$$

and
$$P(i) = \frac{1}{Z} \exp(-1/kT \sum_l n_l^i \epsilon_l), \quad (2a)$$

and the average number of particles in a state l averaged over the ensemble comes to

$$\bar{n}_l = \sum_{\text{all STATES}} n_l^i P(E_i), \quad (4)$$

the sum taken over all STATES of the assembly. Hence, using (2a),

$$\bar{n}_l = - \frac{kT}{Z} \cdot \frac{\partial Z}{\partial \epsilon_l}, \quad (4a)$$

Z may be split up into terms according to the number of particles in a particular state. Then

$$Z = \sum_{N,0}^l + \sum_{N,1}^l + \dots + \sum_{N,N}^l, \quad (5)$$

and from (2) and (4)

$$\bar{n}_l = \frac{1}{Z} \left(\sum_{N,1}^l + 2 \sum_{N,2}^l + 3 \sum_{N,3}^l + \dots + N \sum_{N,N}^l \right), \quad (6)$$

where the subscript N of Z , \sum , \bar{n}_l is a reminder that there are N particles in the assembly, and $\sum_{N,k}^l$ denotes the sum of all those terms in Z for which $n_i^i = k$, for all i .

§ 3.

We wish now to express \bar{n}_l in terms of ϵ_l for different assumptions about the definition of the STATE of the assembly, valid for large values of N . The restriction to large values of N may be put into the form

$$\bar{n}_l = \bar{n}_l (1 + \delta); \quad \delta \rightarrow 0 \text{ for } N \rightarrow \infty. \quad (7)$$

* "STATE" refers to the assembly, "state" to individual particles.

(a) Boltzmann Statistics.

Let Z_N be known for an assembly of N distinguishable particles. If another new particle is added to the assembly a STATE of the assembly of $N+1$ particles arises by bringing this particle into any state r , for any STATE of the assembly of N particles. Hence with $Z = \sum_1 \exp(-\epsilon_r/kT)$

$$Z_{N+1} = Z Z_1$$

because each old STATE gives rise to a set of new STATES, the corresponding terms of Z being multiplied by the same factor Z_1 . Hence by induction

$$Z_N = (Z_1)^N,$$

and from (4 a)

$$\bar{n}_i = \frac{N}{Z} \exp(-\epsilon_i/kT)$$

for the average number of particles found in a particular state of energy ϵ_i . Equation (7) leads in this case to

$$\delta = 1/N.$$

(b) Bose-Einstein Statistics.

Here a STATE i of the assembly is entirely defined by a set of values n_i^i . According to (5)

$$Z_N = \sum_{N,0}^i + \sum_{N,1}^i + \sum_{N,2}^i + \dots + \sum_{N,N}^i, \quad \dots \dots \dots (5)$$

$$Z_{N+1} = \sum_{N+1,0}^i + \sum_{N+1,1}^i + \sum_{N+1,2}^i + \dots + \sum_{N+1,N+1}^i \quad \dots (5a)$$

Whenever the particles are indistinguishable

$$\exp(-\epsilon_i/kT) \sum_{N,a}^i = \sum_{N+1,a+1}^i, \quad \dots \dots \dots (8)$$

for the R.H.S. will contain all terms and only terms with $n_i^i = a+1$ (all i), and $\sum n_i^i$ (for all i) has increased by one. Hence by multiplying (5) by $\exp(-\epsilon_i/kT)$ and using (8) and (5 a)

$$\exp(-\epsilon_i/kT) Z_N = Z_{N+1} - \sum_{N+1,0}^i \quad \dots \dots \dots (9)$$

On differentiating (9) with respect to ϵ_i the second term R.H.S. being independent of ϵ_i vanishes so that

$$-\frac{1}{kT} \exp(-\epsilon_i/kT) Z_N + \exp(-\epsilon_i/kT) \frac{\partial Z_N}{\partial \epsilon_i} = \frac{\partial Z_{N+1}}{\partial \epsilon_i}.$$

Dividing by Z_N and using (4 a)

$$-\frac{1}{kT} \exp(-\epsilon_l/kT) - \frac{1}{kT} \exp(-\epsilon_l/kT) \bar{n}_l = \frac{Z_{N+1}}{Z_N} \frac{1}{Z_{N+1}} \frac{\partial Z_{N+1}}{\partial \epsilon_l} = -\frac{1}{kT} \frac{Z_{N+1}}{Z_N} \bar{n}_l.$$

Inserting (7), the terms containing δ can be neglected so that

$$\bar{n}_l = \frac{1}{A \exp(\epsilon_l/kT) - 1},$$

where $A = Z_{N+1}/Z_N$ is independent of ϵ_l and is evaluated through

$$\sum_l \bar{n}_l = N.$$

(c) *Fermi Statistics.*

n_l^i takes the values 0 or 1. Hence of (5) and (5 a) only two terms remain, and with (8)

$$Z_N = \sum_{N,0}^l + \exp(-\epsilon_l/kT) \sum_{N-1,0}^l, \quad \dots \quad (5b)$$

$$Z_{N+1} = \sum_{N+1,0}^l + \exp(-\epsilon_l/kT) \sum_{N,0}^l. \quad \dots \quad (5c)$$

Combining (5 b) and (5 c)

$$Z_{N+1} = \sum_{N+1,0}^l + \exp(-\epsilon_l/kT) \frac{Z_N - \exp(-2\epsilon_l/kT) \sum_{N-1,0}^l}{N}, \quad \dots \quad (10)$$

and on differentiating (10) with respect to ϵ_l and then dividing by $Z_N/kT \exp(-\epsilon_l/kT)$

$$\frac{kT}{Z_N} \frac{\partial Z_{N+1}}{\partial \epsilon_l} \exp(\epsilon_l/kT) = \frac{kT}{Z_N} \frac{\partial Z_N}{\partial \epsilon_l} - 1 + 2 \exp(-\epsilon_l/kT) \frac{\sum_{N-1,0}^l}{Z_N}.$$

Hence by (4 a), (8) and (6), [as (6) is here reduced to one term]

$$-\frac{Z_{N+1}}{Z_N} \exp(\epsilon_l/kT) \bar{n}_l = -\bar{n}_l - 1 + 2\bar{n}_l,$$

or, using (7), the terms with δ can be neglected so that

$$\bar{n}_l = \frac{1}{A \exp(\epsilon_l/kT) + 1},$$

where $A = Z_{N+1}/Z_N$ is independent of the choice of ϵ_l and is evaluated through $\sum_l \bar{n}_l = N$.

(d) *Photons.*

This differs from case (b) in that N is not conserved. The subscript N in (5) is to be left out. Hence from (5) and (8)

$$Z = \sum_{,0}^l + \exp(-\epsilon_l/kT) \sum_{,0}^l + \exp(-2\epsilon_l/kT) \sum_{,2}^l + \dots$$

$$= \frac{\sum_{,0}^l}{1 - \exp(\epsilon_l/kT)},$$

and using (4 a)

$$\bar{n}_l = \frac{1}{\exp(\epsilon_l/kT) - 1}.$$

It is seen that in the derivations given no other approximation is used but $\bar{n}_l = \bar{n}_l(1 + \delta)$ which is implicitly used also in the method of steepest descent.

REFERENCES.

- DARWIN, C. G., and FOWLER, R. H., 1922, *Phil. Mag.*, **44**, 450.
 FOWLER, R. H., 1936, *Statistical Mechanics* (Cambridge: University Press).
 SCHRÖDINGER, E., 1946, *Statistical Thermodynamics* (Cambridge: University Press).
 TOLMAN, R. C., 1938, *The Principles of Statistical Mechanics* (Oxford: University Press).

LIX. *Investigation of Soft Radiations by Proportional Counter.*—
 IV. *The Beta Spectrum of Ni⁶³.*

By H. W. WILSON and S. C. CURRAN,

Department of Natural Philosophy of the University of Glasgow*.

[Received March 21, 1949.]

THE isotope Ni⁵⁹ is listed in *Science* (1946)† and in various publications (Rosenfeld 1948) as a positron emitter of upper energy limit 50 Kev. The quoted life-time of 15 years suggests that the transition is allowed. The shape of such an allowed positron spectrum was considered of special importance and the proportional counter method as applied by Curran, Angus and Cockroft (1948–49) was obviously well suited to the investigation even if only small quantities of nickel in the form of carbonyl vapour of low specific activity were available. The present work failed to confirm the existence of such a positron emitter but revealed the existence of a soft β -source which can be identified as Ni⁶³.

* Communicated by the Authors.

† Isotope Committee, 1946, *Science*, 103, 697.

A quantity of metallic nickel was irradiated for a few months in a pile* and allowed to stand for about six months. The main activity of the source seemed to be that of Co^{60} but careful chemical separation was carried out by the following method:—

The nickel was converted to chloride, cobalt chloride added as carrier, and precipitation of all cobalt in solution effected by adding α -nitroso- β -naphthol solution. The filtrate was diluted, ammoniated, and the nickel precipitated by dimethylglyoxime. Then the precipitate was dried, oxidized and dissolved in nitric acid. After evaporation of the excess acid, ammonium hydroxide was used to precipitate the nickel as nickel hydroxide. The hydroxide (mixed with 1 per cent mercuric oxide as catalyst) was heated in a stream of hydrogen at a temperature not exceeding 250°C . The active nickel so produced was converted to nickel carbonyl by the action of carbon monoxide at 80°C . This series of chemical reactions was considered adequate to ensure complete separation of the nickel from any contaminant.

The carbonyl so obtained was found to have a very high electron affinity and when used in the proportional tube at a partial pressure of about 1 cm. of mercury the degree of proportionality was unsatisfactory.

Pending the preparation of sources of higher specific activity permitting use at a very reduced pressure, the carbonyl was decomposed by heating and a very thin layer of nickel was deposited on an aluminium foil. Most of the results given here were obtained with this source. The activity of the deposit (spread over an area of 6 in. \times 3 in.) was $\sim 50,000$ disintegrations per minute. Another small source, nickel deposited on wire, was occasionally employed.

The following experiments seemed to indicate definitely that the nickel was not positron-active:—

- (1) Magnetic deflection of the β -particles
 - (a) into an open ended Geiger tube
 - (b) on to a photographic plate

showed that the charge of the particles was negative.

(2) Search for γ -radiation indicated that less than 1 per cent of the disintegrations gave rise to quanta. The source was wrapped round a Geiger tube with a lead cathode (0.4 mm. thick) and the measurements gave:—

Rate with source = 149.1 ± 2.1 counts/min.

Rate without source = 152.1 ± 2.1 counts/min.

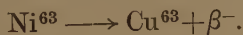
Now two quantum annihilation of the particles, assuming they were all β^+ , would have given an additional 500 counts per minute approximately, as the solid angle was approximately 2π and the efficiency of

* The source was irradiated at A.E.R.E., Harwell. The authors are indebted to the Ministry of Supply and to the Director, Sir J. D. Cockcroft.

the counter could be assumed to be 1 per cent for such quanta. The result implies that the particles were more than 99 per cent negative. It is to be remembered that the specific activity of the source was very small and none of the obvious methods of testing for the sign of the particles could be adopted.

The absorption curve of the negative β -particles was determined by placing a source within a counter and covering it with increasing thickness of aluminium. The range was found to be ~ 6.0 mgm./cm.² indicating an upper energy limit of 60 to 70 Kev.

Examination of a chart of known stable and radioactive isotopes in the neighbourhood of nickel shows that Ni^{59} should decay by positron activity or by K-capture. No evidence of positrons, KX-radiations or Auger electrons could be detected in our experiments. Indeed a separate careful search by the proportional method for any soft electromagnetic radiation of energy between 2 and 60 Kev. failed to reveal its presence, establishing the fact that the spectrum was simple. It was established that quanta in this energy range were emitted in less than 1 per cent of the disintegrations. Hence we are forced to conclude that the amount of Ni^{59} obtained by pile irradiation of nickel is negligibly small. Moreover, it has been shown recently (Swartout 1946, Conn 1946), that the known short period β activity (2.6 hrs.) of nickel should be definitely ascribed to Ni^{65} . On the other hand, the relatively rare stable isotype Ni^{62} (3.8 per cent abundant) can yield Ni^{63} which is likely to decay by a negative β -process, *i.e.*

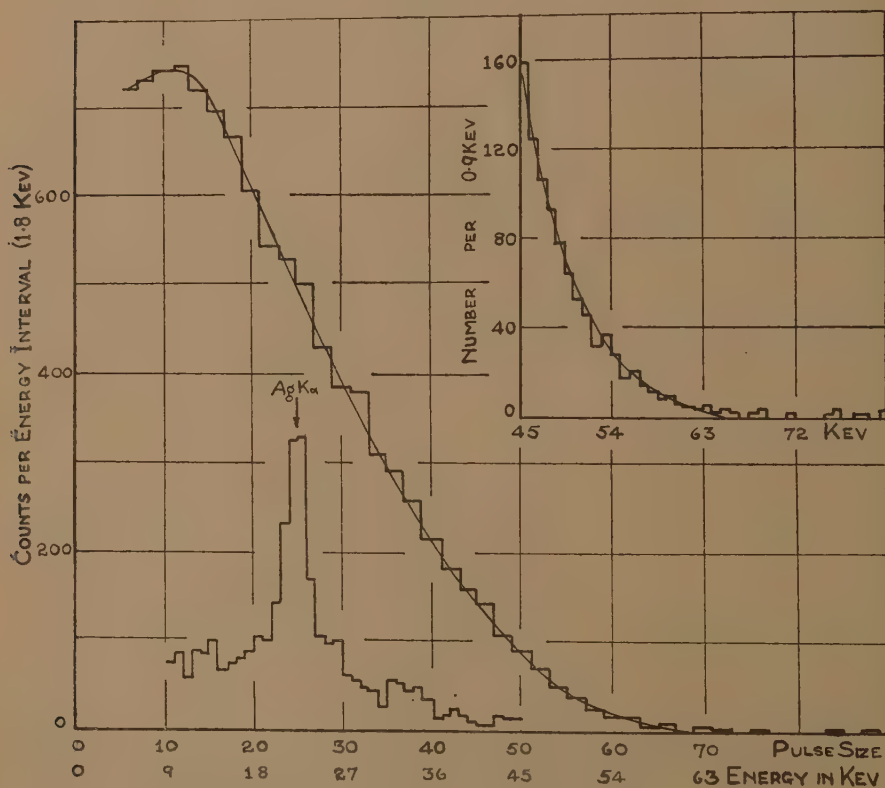


Thus Ni^{63} appears to be the only possible nickel isotype with which the activity discussed here can be associated. The half-life of Ni^{59} has been given as 15 years. Our present observations indicate the Ni^{63} will have a life of at least this magnitude. We are now determining the half-life of the isotope Ni^{63} with some precision. It would appear that the specially interesting positron activity ascribed in previous work to Ni^{59} may unfortunately be spurious. It is possible that the β^+ activity associated formerly with Ni^{59} may in fact have been the soft β -emission of Ni^{63} . The form of this β -spectrum was examined by the proportional tube technique. The aluminium sheet bearing the very thin active deposit was placed in a proportional tube (diam. $2\frac{1}{2}$ in.) which was filled with argon (90 per cent) and methane (10 per cent) to a pressure of two atmospheres. The active length of the proportional counter was 7 in. and the foil on which the source was deposited took the form of a cylinder pressing against the wall of the counter and situated in the middle of the tube.

Thus the gas gain was constant over the entire volume in which the particles were emitted. It was not necessary, therefore, to consider the effect produced by variation of gain in the vicinity of the ends of the wire (Curran and Cockroft 1949). The histogram of the pulses giving

the β -spectrum is shown in fig. 1. The inset figure (fig. 1), together with the histogram of the calibrating radiation ($K\alpha$ X-radiation of silver of energy 22.07 Kev. superposed on the spectrum) gives the value of the end-point energy, viz. $E_{\max} = 63 \pm 2$ Kev. The maximum intensity occurs at an energy of about 10 Kev. The form of the spectrum at very low energy values is shown continuous with the main part of fig. 1 although it was examined at a higher value of gas gain. It is in this region that modification of the true form by reflection at the aluminium support gives rise to some uncertainty.

Fig. 1.



β^- spectrum of Ni^{63} with calibration and end-point (inset).

In fig. 2, we have plotted $(N/pWG)^{\frac{1}{2}}$ against the kinetic energy of the β -particles. N is measured intensity per energy interval, p the momentum and W the total energy (including rest energy). The influence of the Coulomb field of the nucleus of medium charge Z is taken into account in the factor G where $G(W, Z)$ is given by the equation

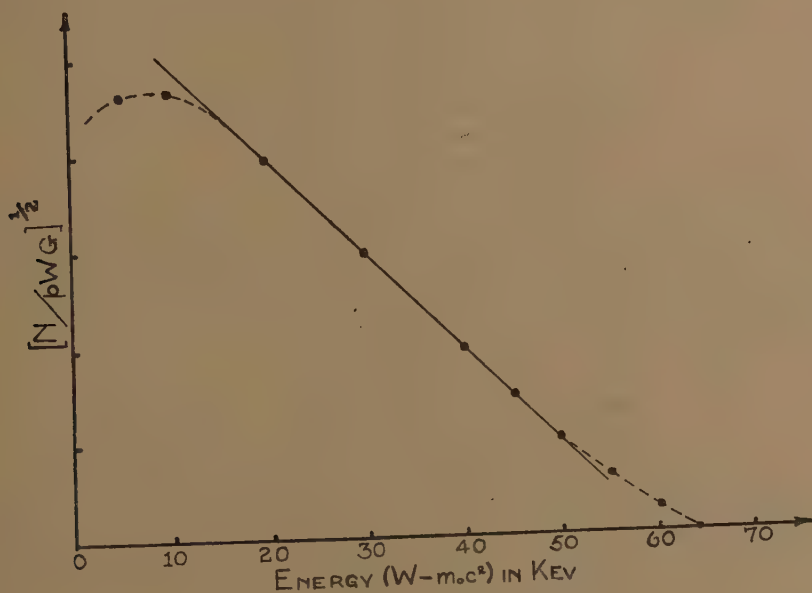
$$G(W, Z) = \text{const.} \times p^{2s-2} \phi / (e^{-\phi} - 1)$$

where $s = \sqrt{1 - \alpha^2 Z^2}$, $\alpha = 1/137$, $\phi = 2\pi Z/p$ and $\beta = v/c$.

The curve approximates rather closely to a straight line down to 15 Kev. Beyond this the intensity N seems to drop below that expected for an allowed transition and the results suggest that the transition may be first forbidden. However, reflection at the source support makes this conclusion somewhat tentative. Above 50 Kev. there is evidence of a slight excess of particles.

Since completing the work described above, we have seen that private communications of Friedlander and Goldhaber and Brosi and Conn to Seaborg and Perlman (1948) define Ni^{63} as a β^- source of long life (~ 300 yr.). For the measured upper energy limit E_0 of $1.123 m_0c^2$ given here the $tF(E_0)$ value obtained with $t=300$ yr. rather definitely

Fig. 2.



Fermi plot of spectrum.

suggests that the transition is first forbidden, but a half-life t of 15 to 30 yr. would be more or less consistent with an allowed transition. The proposed direct measurement of t and more accurate examination of the spectrum in the region of very low energy may resolve this question completely.

It is hoped that a more thorough check on Fermi theory for this β -emitter can be made later by using the source in the gaseous form— $\text{Ni}(\text{CO})_4$ —at a very low partial pressure. The source is of particular interest since the K-shell energy of the emitting atom (cobalt) is of the same order as the average energy of the β -particles. A study of the spectrum in greater detail may lead to interesting results regarding screening and other extra-nuclear effects on the β -emission.

We wish to thank Professor P. I. Dee, F.R.S., for encouraging the prosecution of the work, and Mr. B. Touschek for helpful discussions on certain theoretical aspects.

REFERENCES.

- CONN, E. E. et al, 1946, *Phys. Rev.*, **70**, 768.
 CURRAN, S. C., ANGUS, J., and COCKROFT, A. L., 1948, *Nature, Lond.*, **162**, 302;
 1949, *Phil. Mag.*, **40**, 36.
 KONOPINSKI, E. J., 1943, *Rev. Mod. Phys.*, **15**, 209.
 ROSENFELD, L., "Nuclear Forces", 1948, Vol. 1, p. 506, North-Holland
 Publishing Co., Amsterdam.
 SEABORG, G. T. and PERLMAN, I., 1948, *Rev., Mod. Phys.*, **20**, 585.
 SWARTOUT, J. A. et al., 1946, *Phys. Rev.*, **70**, 323.

LX. *On the Motion of a Vector Meson in a Homogeneous Magnetic Field.*

By N. SYMONDS,

Dublin Institute for Advanced Studies *.

[Received March 22, 1949.]

SUMMARY.

The Kemmer equation is applied to find exact solutions for the wave-functions of a vector meson in a homogeneous magnetic field. The method used is then generalized and leads to a new form for the second-order Kemmer equation in the presence of an electromagnetic field.

§ 1. INTRODUCTION.

THE problem of finding wave-functions describing the motion of a scalar particle in a homogeneous magnetic field has been solved by Plesset (1930), and the solutions for an electron have been given by Huff (1931). It is the purpose of this paper to determine the solutions when the particle concerned is a vector meson.

The formulation of meson theory used is that due to Kemmer (1939), and two methods of treating the Kemmer equation are given. The first, which is worked out in detail, involves using the special coordinate system for which the momentum of the meson in the direction of the magnetic field is zero. This reduces the number of matrices occurring in the Kemmer equation to three, and a representation can then be used in which these three matrices are reducible.

The second method is based on some properties of the Kemmer matrices due to Schrödinger (1943), and by a straightforward generalization this leads to a new form for the second-order Kemmer equation in the presence of an electromagnetic field, which may be useful in subsequent meson calculations.

* Communicated by Professor W. Heitler, F.R.S.

§ 2. FUNDAMENTAL EQUATIONS.

In the presence of an electromagnetic potential ϕ_i ($i=1, 2, 3, 4$), the Kemmer equation is

$$\beta_i D_i \psi + \kappa \psi = 0, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where *

$$D_i = \frac{\partial}{\partial x^i} - ie\phi_i,$$

and the β_i are square matrices of degree ten satisfying the commutation relations

$$\beta_i \beta_j \beta_k + \beta_k \beta_j \beta_i = \delta_{ij} \beta_k + \delta_{jk} \beta_i. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

On introducing the electromagnetic field strength F_{ij} , and matrices I_{ij} defined by

$$I_{ij} = \beta_i \beta_j - \beta_j \beta_i, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

the second-order equation resulting from (1) becomes

$$(D_k D_k - \kappa^2) \psi = -\frac{ie}{2\kappa} D_k (\beta_i \beta_k \beta_j - \delta_{ik} \beta_j) F_{ij} \psi + \frac{ie}{2} F_{ij} I_{ij} \psi. \quad . \quad . \quad (4)$$

The particular case of a homogeneous magnetic field in the z -direction can be described by taking

$$\phi_2 = Hx, \quad \phi_1 = \phi_3 = \phi_4 = 0,$$

so

$$F_{12} = -F_{21} = H.$$

The specific equation to be solved is therefore

$$\left[\beta_1 \frac{\partial}{\partial x} + \beta_2 \left(\frac{\partial}{\partial y} - ieHx \right) + \beta_3 \frac{\partial}{\partial z} - i\beta_4 \frac{\partial}{\partial t} \right] \psi + \kappa \psi = 0, \quad . \quad . \quad (5)$$

and we seek for solutions such that

$$\frac{\partial}{\partial y} = ip_y, \quad \frac{\partial}{\partial z} = ip_z, \quad \frac{\partial}{\partial t} = -iE. \quad . \quad . \quad . \quad . \quad (6)$$

The method we shall use to obtain these solutions is to work directly with the second-order equation (4), using the properties of the β matrices. We now proceed to show how this can be done relatively simply by using the special coordinate system in which $p_z = 0$.

§ 3. THE SPECIAL COORDINATE SYSTEM.

The use of the special system with $p_z = 0$ involves no loss of generality as it is always possible to pass by a Lorentz transformation from this system to the general one in which $p_z \neq 0$. The relevant formulæ for this type of transformation have been given by Schrödinger (1943), and if we

* Throughout we use atomic units with $\hbar = c = 1$.

On choosing β_1 , β_2 and β_4 to be as given in (10), the three representations given there can be treated quite independently, and we now proceed to solve for each in turn the second-order equation (4), which for this problem is

$$(\mathbf{D}_k \mathbf{D}_k - \kappa^2)\psi = -\frac{ie\mathbf{H}}{\kappa}(-\beta_2\beta_1^2\mathbf{D}_1 + \beta_1\beta_2^2\mathbf{D}_2 + \beta_1\beta_4\beta_2\mathbf{D}_4)\psi + ie\mathbf{H}\mathbf{I}_{12}\psi. \quad (11)$$

§ 4. THE REPRESENTATION $\mathbf{R}_4\{1, 0\}$.

The matrices of this representation possess the property

$$\beta_1\beta_4\beta_2 = \beta_1\beta_2\beta_4 = \beta_2\beta_1\beta_4 = 0. \quad (12)$$

We take a representation in which \mathbf{I}_{12} is diagonal. This is

$$\beta_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & -i \\ \cdot & \cdot & \cdot & \cdot \\ 1 & i & \cdot & \cdot \end{pmatrix} \quad \beta_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} \cdot & \cdot & \cdot & -i \\ \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & \cdot & \cdot \\ i & 1 & \cdot & \cdot \end{pmatrix} \quad \beta_4 = \begin{pmatrix} \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & 1 \\ \cdot & \cdot & 1 & \cdot \end{pmatrix}, \quad (13)$$

where the dots all represent zeros.

Using the commutation relations (2) it is found that the matrices occurring on the right side of equation (11) have the values

$$-\beta_2\beta_1^2 = \mathbf{I}_{12}\beta_1,$$

$$\beta_1\beta_2^2 = \mathbf{I}_{12}\beta_2,$$

$$\beta_1\beta_4\beta_2 = 0 = \mathbf{I}_{12}\beta_4,$$

from (12), so

$$(-\beta_2\beta_1^2\mathbf{D}_1 + \beta_1\beta_2^2\mathbf{D}_2 + \beta_1\beta_4\beta_2\mathbf{D}_4)\psi = \mathbf{I}_{12}(\beta_i\mathbf{D}_i)\psi = -\kappa\mathbf{I}_{12}\psi.$$

The second-order equation therefore becomes

$$(\mathbf{D}_k \mathbf{D}_k - \kappa^2)\psi = 2ie\mathbf{H}\mathbf{I}_{12}\psi, \quad (14)$$

and on changing the independent variable by the substitution

$$\xi = (e\mathbf{H}x - p_y)/(e\mathbf{H})^{\frac{1}{2}}, \quad (15)$$

the equation (14) becomes

$$\frac{d^2\psi}{d\xi^2} + \left(\frac{\mathbf{E}^2 - \kappa^2}{e\mathbf{H}} - 2i\mathbf{I}_{12} - \xi^2 \right)\psi = 0. \quad (16)$$

The solutions to this equation are in terms of the Hermitian polynomials $\mathbf{H}_n(\xi)$, and using the representation (13), it is found that the Kemmer equation has the following solution

$$\psi = b \begin{vmatrix} (e\mathbf{H})^{\frac{1}{2}}v_{n+1} \\ in(e\mathbf{H})^{\frac{1}{2}}v_{n-1} \\ \mathbf{E}v_n \\ \kappa v_n \end{vmatrix}, \quad (17)$$

2 2 2

where

$$v_n = (eH)^{\frac{1}{2}} \pi^{-\frac{1}{2}} 2^{-n/2} e^{-\xi^2/2} H_n(\xi), \quad n \geq 0 \quad (18)$$

and

$$E^2 - \kappa^2 = eH(2n+1). \quad (19)$$

The constant b must be determined by normalization, and there are two cases to consider (Heitler 1943) corresponding to positive and negative charge. We find

$$\left. \begin{array}{ll} \text{positive charge} & b^{-2} = n! 2E\kappa, \\ \text{negative charge} & b^{-2} = -n! 2E\kappa. \end{array} \right\} \quad (20)$$

The positive energy values therefore correspond to positive charge, while the negative energy values correspond to negative charge. As pointed out by Heitler, this state of affairs is an integral part of meson theory, and is essential in order that measurable values should have a unique interpretation.

The solution given above is the same as that for a scalar particle. This is related to the fact that the spin operator for a vector meson has zero for one of its eigen-values.

§ 5. THE REPRESENTATION $R_4\{1, 1\}$.

This representation has the properties

$$\beta_1^2 + \beta_2^2 + \beta_4^2 = 2, \quad (21)$$

$$\beta_4 = -iI_{12}, \quad \beta_1 = -iI_{24}, \quad \beta_2 = -iI_{41}. \quad (22)$$

A representation with I_{12} diagonal is

$$\beta_1 = \frac{1}{\sqrt{2}} \begin{vmatrix} \cdot & \cdot & 1 \\ \cdot & \cdot & 1 \\ 1 & 1 & \cdot \end{vmatrix}, \quad \beta_2 = \frac{1}{\sqrt{2}} \begin{vmatrix} \cdot & \cdot & -i \\ \cdot & \cdot & i \\ i & -i & \cdot \end{vmatrix}, \quad \beta_4 = \begin{vmatrix} 1 & \cdot & \cdot \\ \cdot & -1 & \cdot \\ \cdot & \cdot & \cdot \end{vmatrix}. \quad (23)$$

The matrix terms occurring in equation (11) now have the values

$$-\beta_2\beta_1^2 = I_{12}\beta_1,$$

$$\beta_1\beta_2^2 = I_{12}\beta_2,$$

$$\beta_1\beta_4\beta_2 = \beta_1I_{42} + I_{12}\beta_4 + \beta_2I_{14} + \beta_2\beta_4\beta_1,$$

from (3), so

$$\begin{aligned} \beta_1\beta_4\beta_2 - \beta_2\beta_4\beta_1 &= 2\beta_1\beta_4\beta_2 \\ &= I_{12}\beta_4 + \beta_1I_{42} + \beta_2I_{14} \\ &= 2I_{12}\beta_4 - 2i, \end{aligned}$$

using (21) and (22).

Inserting these expressions into equation (11) we find the following second-order equation

$$(D_k D_k - \kappa^2)\psi = eH(2iI_{12} + E/\kappa)\psi,$$

and on changing the variable by the substitution (15) this becomes

$$\frac{d^2\psi}{d\xi^2} + \left(\frac{E^2 - \kappa^2}{eH} - \frac{E}{\kappa} - 2iI_{12} - \xi^2 \right) \psi = 0. \quad (24)$$

The solutions are therefore again in terms of Hermitian polynomials, and with the representation (23) the following solution to the Kemmer equation is obtained, valid when $n \geq 2$,

$$\psi = b \begin{vmatrix} (E + \kappa)v_n \\ (n-1)(E - \kappa)v_{n-2} \\ -(E^2 - \kappa^2)(eH)^{-\frac{1}{2}}v_{n-1} \end{vmatrix}, \quad (25)$$

where

$$E^2 - \kappa^2 = eH(2n - 1 + E/\kappa) \quad (26)$$

and

$$b^{-2} = \pm (n-1)! [(E - \kappa)^2 + 4nE\kappa]. \quad (27)$$

The plus or minus sign in (27) refers to positive and negative charge respectively, and it can be again seen that positive energies correspond to positive charge and negative energies to negative charge.

Special attention has to be paid to the cases when $n=0$ or 1. By studying the first-order equations it is found that the only energy-values leading to physically admissible solutions are $n=0$, $E=\kappa$ and $n=1$, $E=\kappa + eH/\kappa$. The associated solutions in these cases are given in §7. Recently Luttinger (1948) has shown that the solution with $E=\kappa$ in the electron case can be used to calculate the radiative correction to the electron magnetic moment, and the existence of a similar solution here therefore admits the possibility of his method being applied in the case of the meson magnetic moment also.

§6. THE REPRESENTATION $R_4^*\{1, 1\}$.

This representation has the properties

$$\beta_1^2 + \beta_2^2 + \beta_4^2 = 2,$$

$$\beta_4 = iI_{12}, \quad \beta_1 = iI_{24}, \quad \beta_2 = iI_{41},$$

and a representation with I_{12} diagonal is

$$\beta_1 = \frac{1}{\sqrt{2}} \begin{vmatrix} \cdot & \cdot & 1 \\ \cdot & \cdot & 1 \\ 1 & 1 & \cdot \end{vmatrix} \quad \beta_2 = \frac{1}{\sqrt{2}} \begin{vmatrix} \cdot & \cdot & -i \\ \cdot & \cdot & i \\ i & -i & \cdot \end{vmatrix} \quad \beta_4 = \begin{vmatrix} -1 & \cdot & \cdot \\ \cdot & 1 & \cdot \\ \cdot & \cdot & \cdot \end{vmatrix}. \quad (28)$$

Proceeding exactly as in the previous section we obtain the second-order equation

$$\frac{d^2\psi}{d\xi^2} + \left(\frac{E^2 - \kappa^2}{eH} + \frac{E}{\kappa} - 2iI_{12} - \xi^2 \right) \psi = 0. \quad (29)$$

Comparing this equation with (24), it can be seen that the only difference between the two cases lies in the sign of E . This means that the positive energy solutions in §5 correspond to the negative energy solutions here and vice versa, and it also follows that taken together these two sets of solutions are completely symmetric in positive and negative energy.

The explicit solutions for this representation are given in the next section.

§7. SUMMARY OF SOLUTIONS.

We here collect together the various solutions found in the previous sections :

$$\begin{aligned} \text{I. } E^2 - \kappa^2 &= eH(2n+1), \\ b^{-2} &= \pm n! 2E\kappa, \\ n &\geq 0, \end{aligned} \quad \psi = b \begin{vmatrix} (eH)^{\frac{1}{2}} v_{n+1} \\ in(eH)^{\frac{1}{2}} v_{n-1} \\ Ev_n \\ \kappa v_n \end{vmatrix}.$$

$$\begin{aligned} \text{II. } E^2 - \kappa^2 &= eH(2n-1+E/\kappa), \\ b^{-2} &= \pm (n-1)! [(E-\kappa)^2 + 4nE\kappa], \\ n &\geq 2, \end{aligned} \quad \psi = b \begin{vmatrix} (E+\kappa)v_n \\ (n-1)(E-\kappa)v_{n-2} \\ -(E^2-\kappa^2)(eH)^{-\frac{1}{2}}v_{n-1} \end{vmatrix}.$$

$$\begin{aligned} n=0, \\ E=\kappa, \end{aligned} \quad \psi = \begin{vmatrix} v_0 \\ 0 \\ 0 \end{vmatrix}, \quad \begin{aligned} n=1 \\ E=\kappa+eH/\kappa \end{aligned} \quad \psi = \begin{vmatrix} v_1 \\ 0 \\ -(eH)^{\frac{1}{2}}\kappa^{-1}v_0 \end{vmatrix}.$$

$$\begin{aligned} \text{III. } E^2 - \kappa^2 &= eH(2n-1-E/\kappa), \\ b^{-2} &= \pm (n-1)! [4nE\kappa - (E+\kappa)^2], \\ n &\geq 2, \end{aligned} \quad \psi = b \begin{vmatrix} (E-\kappa)v_n \\ (n-1)(E+\kappa)v_{n-2} \\ (E^2-\kappa^2)(eH)^{-\frac{1}{2}}v_{n-1} \end{vmatrix}.$$

$$\begin{aligned} n=0, \\ E=-\kappa, \end{aligned} \quad \psi = \begin{vmatrix} v_0 \\ 0 \\ 0 \end{vmatrix}, \quad \begin{aligned} n=1, \\ E=-\kappa-eH/\kappa, \end{aligned} \quad \psi = \begin{vmatrix} v_1 \\ 0 \\ -(eH)^{\frac{1}{2}}\kappa^{-1}v_0 \end{vmatrix}.$$

where $v_n = (eH)^{\frac{1}{2}} \pi^{-\frac{1}{2}} 2^{-n/2} e^{-\xi^2/2} H_n(\xi)$, $\xi = (eHx - p_y)/(eH)^{\frac{1}{2}}$.

§8. ALTERNATIVE METHOD OF SOLUTION.

Schrödinger has shown that from the four 10-rowed matrices $\beta_1, \beta_2, \beta_3$ and β_4 it is possible to find a fifth matrix β_0 say, which satisfies the same commutation relations as the original β matrices. That is $\beta_0, \beta_1, \beta_2, \beta_3$ and β_4 are said to form a pentad. If we introduce a further suffix 5, it is possible to write

$$\beta_0 = iI_{05}, \quad \beta_1 = iI_{15}, \quad \beta_2 = iI_{25}, \quad \beta_3 = iI_{35}, \quad \beta_4 = iI_{45}. \quad \dots \quad (30)$$

Together with the ten matrices defined in the usual way, there are therefore fifteen matrices I_{ab} ($a, b=0, 1, 2, 3, 4, 5$), and associated with each of these are defined the quantities

$$\eta_{ab}=1+2I_{ab}^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

The matrices η_{ab} and I_{ab} possess, in particular, the properties

$$(a) \quad \eta_{ab}I_{bc}=-\eta_{ac}I_{bc}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

$$(b) \quad I_{ab}I_{cd}=I_{cd}I_{ab}=-\frac{i}{2}(1-\eta_{ab})I_{ef}, \quad . \quad . \quad . \quad . \quad (33)$$

(c) the sum of the η s in a pentad equals -1 , *e. g.*

$$\eta_{05}+\eta_{15}+\eta_{25}+\eta_{35}+\eta_{45}=-1. \quad . \quad . \quad . \quad . \quad (34)$$

where a, b, c, d, e, f are different suffixes forming an even permutation of $0, 1, 2, 3, 4, 5$.

In the general case the second-order equation to be solved is, from (4),

$$(D_k D_k - \kappa^2)\psi = -ieH/\kappa \{-\beta_2\beta_1^2 D_1 + \beta_1\beta_2^2 D_2 + \beta_1\beta_3\beta_2 D_3 + \beta_1\beta_4\beta_2 D_4\}\psi + ieHI_{12}\psi. \quad . \quad . \quad . \quad (35)$$

The matrices occurring in the large bracket have the values

$$-\beta_2\beta_1^2 = I_{12}\beta_1,$$

$$\beta_1\beta_2^2 = I_{12}\beta_2,$$

$$2\beta_1\beta_3\beta_2 = \beta_1 I_{32} + I_{12}\beta_3 + \beta_2 I_{13},$$

from (3), but

$$\begin{aligned} \beta_1 I_{32} + \beta_2 I_{13} &= \frac{1}{2}(2 - \eta_{15} - \eta_{25})I_{04} \quad \text{using (30) and (33)} \\ &= \frac{1}{2}(\eta_{35} - 1)I_{04} + \frac{1}{2}(\eta_{05} + \eta_{45})I_{04} + 2I_{04} \quad (\text{from (34)}) \\ &= I_{12}\beta_3 + 2I_{04} \quad \text{using (32) and (33),} \end{aligned}$$

so

$$\beta_1\beta_3\beta_2 = I_{12}\beta_3 + I_{04},$$

and similarly

$$\beta_1\beta_4\beta_2 = I_{12}\beta_4 - I_{03}.$$

Equation (35) therefore becomes

$$(D_k D_k - \kappa^2)\psi = 2ieHI_{12}\psi + (eH/\kappa)(p_z I_{04} - i\bar{E}I_{03})\psi, \quad . \quad . \quad (36)$$

where we again denote the energy in the system with $p_z \neq 0$ by \bar{E} . It can now be seen that the operator $(p_z I_{04} - i\bar{E}I_{03})$ commutes with all terms in equation (36), and so can be given its eigen-values which are

$$0, \quad \pm(\bar{E}^2 - p_z^2)^{\frac{1}{2}} = 0, \quad \pm E,$$

from (7).

The following three equations for the determination of the energy-values are therefore contained in equation (36)

$$(D_k D_k - \kappa^2)\psi = 2ieH I_{12}\psi + \frac{eH}{\kappa} \begin{cases} E \\ 0 \\ -E \end{cases} \psi, \quad \dots \quad (37)$$

and it can be verified that these are identical with the three sets of equations solved in the preceding part of the paper.

A straightforward generalization of the procedure in this section then leads to the following form for the second-order Kemmer equation in the presence of an electromagnetic field.

$$(D_k D_k - \kappa^2)\psi = -\frac{ie}{2\kappa} \frac{\partial F_i}{\partial x} (\beta_i \beta_k \beta_j - \delta_{ik} \beta_j)\psi - \frac{ie}{2\kappa} \epsilon_{ijkl} F_{ij} D_k I_{0l}\psi + ie F_{ij} I_{ij}\psi, \quad \dots \quad (38)$$

where ϵ_{ijkl} is the well-known tensor, completely antisymmetric in the indices i, j, k, l .

In conclusion I would like to thank Professor Heitler for suggesting this problem.

APPENDIX.

With the representations of β_1, β_2 and β_4 used in this paper, β_3 is given by

$$\beta_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} \begin{array}{cccc|cccc|cccc} \cdot & \cdot & \cdot & \cdot & i & \cdot & \cdot & \cdot & -i & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & -1 & \cdot & \cdot & \cdot & 1 & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & -i & \cdot & \cdot & \cdot & -i \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array} \\ \hline \begin{array}{cccc|cccc|cccc} -i & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & -1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & i & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array} \\ \hline \begin{array}{cccc|cccc|cccc} i & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & 1 & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & i & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{array} \end{pmatrix}.$$

REFERENCES.

BHABHA, 1945, *Rev. Mod. Phys.*, **17**, 200.
 HEITLER, 1943, *Proc. Roy. Irish Acad. A*, **49**, 16.
 HUFF, 1931, *Phys. Rev.*, **38**, 501.
 KEMMER, 1939, *Proc. Roy. Soc. A*, **173**, 91.
 LUTTINGER, 1948, *Phys. Rev.*, **74**, 893.
 MURNAGHAN, 1938, *Theory of Group Representations*, p. 287.
 PLESSET, 1930, *Phys. Rev.*, **36**, 1728.
 SCHRÖDINGER, 1943, *Proc. Roy. Irish Acad. A*, **48**, 135 ; **49**, 29.

LXI. *On Sommerfeld's "Radiation Condition."*

By F. V. ATKINSON, M.A., D.Phil.*

[Received October 7, 1947.]

§ 1.

It is well-known that in problems of wave-motion involving an infinite medium the equations of wave-motion and boundary conditions do not by themselves determine the solution uniquely. The recognition of this fact appears to be due to Sommerfeld †, who has proposed restrictions on the behaviour of the wave at infinity. In the simplest case, to which I confine myself, the problem of uniqueness may then be formulated as follows :—

It is required to prove that there is no function u such that

- (i) u vanishes at all points on the closed surface S ,
- (ii) u has continuous 2nd derivatives on and outside S ,
- (iii) $\nabla^2 u + k^2 u = 0$ on and outside S ,
- (iv) ru remains bounded as $r \rightarrow \infty$, uniformly in all directions,
- (v) $r \left(iku - \frac{\partial u}{\partial r} \right) \rightarrow 0$ as $r \rightarrow \infty$, also uniformly in all directions.

In the last two conditions r denotes the distance from any fixed point. Condition (iv) is Sommerfeld's "condition of finiteness" (Endlichkeitsbedingung); condition (v) is his "radiation condition" (Ausstrahlungsbedingung).

This proposition is often referred to as an established fact, particularly by writers on diffraction and allied topics, though I have not been able to find any reference to a published proof. The two considerations advanced by Sommerfeld do not, I think, purport to be rigorous arguments ‡. He points out that such a function, satisfying conditions (i)–(v) above, cannot exist if the corresponding Green's function exist; however, the existence of the Green's function seems more difficult to establish than the original proposition. His other argument, a physical one, presupposes a scalar analogue of the energy-flow property of an electromagnetic field. It does not seem easy to validate this suggestion directly, but the following investigation does proceed on related lines.

In this paper I give a proof that Sommerfeld's conditions make the problem unique. I show also that his two conditions may be replaced by one, and that within certain limits their exact form is a matter of indifference.

* Communicated by the Author.

† Sommerfeld (1), pp. 326–334. See also Sommerfeld (2), pp. 803–808, Baker and Copson (3), pp. 25–28.

‡ See the concluding paragraph of Sommerfeld (1), p. 353.

§ 2.

I need the following lemma :—

Lemma. Let u be such that

- (i) $\nabla^2 u + k^2 u = 0$ on and outside the closed surface S ,
- (ii) u has continuous 2nd derivatives on and outside S ,
- (iii) $re^{ikr} \left\{ \left(ik - \frac{1}{r} \right) u - \frac{\partial u}{\partial r} \right\} \rightarrow 0$ as $r \rightarrow \infty$, uniformly in all directions,

where r denotes the distance from any fixed point. Then u can be expanded in a series of the form

$$u = e^{ikr} \sum_{n=1}^{\infty} a_n r^{-n},$$

where the a_n are continuous functions of grad r , the direction of the radius vector, and are independent of r . The series is absolutely and uniformly convergent in a certain region of the form $r > \text{const.}$

Proof: Let P be a point outside S , and let Σ be a sphere centre P , radius R , where R is so large that S lies entirely inside Σ . Then

$$4\pi u_P = \iint_{S+\Sigma} \frac{e^{ikr}}{r} \left\{ \frac{\partial u_Q}{\partial n} - \left(ik - \frac{1}{r} \right) u_Q \frac{\partial r}{\partial n} \right\} d\sigma_Q,$$

where $d\sigma_Q$ is the element of area at the point Q , r is the distance PQ , and the normal is directed inwards on S and outwards on Σ . But the integral over Σ may be written

$$\frac{e^{ikR}}{R} \iint_{\Sigma} \left\{ \frac{\partial u_Q}{\partial R} - \left(ik - \frac{1}{R} \right) u_Q \right\} d\sigma_Q,$$

which $\rightarrow 0$ as $R \rightarrow \infty$, by condition (iii) of the lemma. Hence

$$u_P = \frac{1}{4\pi} \iint_S \frac{e^{ikr}}{r} \left\{ \frac{\partial u_Q}{\partial n} - \left(ik - \frac{1}{r} \right) u_Q \frac{\partial r}{\partial n} \right\} d\sigma_Q, \quad \dots \quad (2.1)$$

for all points P outside S .

Choose now a fixed point X . Let r_0, r_1 denote PX, PQ , respectively, where Q is a variable point on S . Denote r_0^{-1} by x , XQ by ρ , and $\angle PXQ$ by ψ . Then we have

$$\begin{aligned} u_P &= \frac{1}{4\pi} \iint_S \frac{e^{ikr_1}}{r_1} \left\{ \frac{\partial u_Q}{\partial n} - \left(ik - \frac{1}{r_1} \right) u_Q \frac{\partial r_1}{\partial n} \right\} d\sigma_Q, \\ r_1 &= \sqrt{(r_0^2 - 2r_0\rho \cos \psi + \rho^2)}, \\ r_1 - r_0 &= r_0 \left\{ \sqrt{\left(1 - \frac{2\rho \cos \psi}{r_0} + \frac{\rho^2}{r_0^2} \right)} - 1 \right\}, \\ &= \frac{1}{x} \{ \sqrt{(1 - 2x\rho \cos \psi + x^2\rho^2)} - 1 \}, \\ r_0 r_1^{-1} &= (1 - 2x\rho \cos \psi + x^2\rho^2)^{-\frac{1}{2}}. \end{aligned}$$

Hence

$$\frac{e^{ikr_1}}{r_1} \bigg/ \frac{e^{ikr_0}}{r_0} = (1 - 2x\rho \cos \psi + x^2\rho^2)^{-\frac{1}{2}} \exp \frac{ik}{x} \{ \sqrt{(1 - 2x\rho \cos \psi + x^2\rho^2)} - 1 \}.$$

Here the right-hand side is an analytic function of x in the region $|2x\rho \cos \psi - x^2\rho^2| < 1$. It may therefore be expanded in a power series in x in which the coefficients depend only on ρ and ψ , and which will be uniformly convergent at any rate in the region $|x| \leq 1/3\rho$. The function is moreover bounded and continuous in the region $|\rho| < \rho_0$, ψ real, $|x| \leq 1/3\rho_0$, even when x takes complex values. Hence in this ρ , ψ , and x region the power series will be uniformly convergent and the coefficients will be continuous functions of ρ and ψ .

It follows that

$$\left(\frac{e^{ikr_0}}{r_0}\right)^{-1} \iint_S \frac{e^{ikr_1}}{r_1} \frac{\partial u_Q}{\partial n} d\sigma_Q$$

may be expanded in a power series in x , in which the coefficients depend only on the direction $\text{grad } r_0$, and depend on it continuously. The series will be uniformly convergent in the region $r_0 \geq 3\rho_0$, where ρ_0 is the greatest distance of any point of S from X .

The same result holds for

$$\left(\frac{e^{ikr_0}}{r_0}\right)^{-1} \iint_S \frac{e^{ikr_1}}{r_1} \left(ik - \frac{1}{r_1}\right) \frac{\partial r_1}{\partial n} u_Q d\sigma_Q.$$

We have

$$\begin{aligned} ik - \frac{1}{r_1} &= ik - x(1 - 2x\rho \cos \psi + x^2\rho^2)^{-\frac{1}{2}}, \\ \frac{\partial r_1}{\partial n} &= \left(\rho \frac{\partial \rho}{\partial n} - r_0 \cos \psi \frac{\partial \rho}{\partial n} + \rho r_0 \sin \psi \frac{\partial \psi}{\partial n} \right) (r_0^2 - 2r_0\rho \cos \psi + \rho^2)^{-\frac{1}{2}} \\ &= \left(\rho x \frac{\partial \rho}{\partial n} - \cos \psi \frac{\partial \rho}{\partial n} + \rho \sin \psi \frac{\partial \psi}{\partial n} \right) (1 - 2x\rho \cos \psi + x^2\rho^2)^{-\frac{1}{2}}, \end{aligned}$$

so that both these expressions are analytic functions of x in the region $|x| \leq 1/3\rho$. The argument now proceeds as before.

Combining these results with (2.1) we see that

$$\left(\frac{e^{ikr_0}}{r_0}\right)^{-1} u_P,$$

can be expanded as a power series in $1/r_0$, which is uniformly convergent in the region $r_0 \geq 3\rho_0$. This proves the lemma.

§ 3.

A simple consequence of the above lemma is the following result on boundary conditions:—

Theorem I. *Let the imaginary part of k be zero or positive, and let u be such that*

$$(i) \quad \nabla^2 u + k^2 u = 0 \text{ outside and on } S,$$

(ii) u has continuous 2nd derivatives on and outside S .

Then the following boundary conditions at infinity

$$re^{ikr} \left\{ \left(ik - \frac{1}{r} \right) u - \frac{\partial u}{\partial r} \right\} \rightarrow 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (A)$$

$$r^3 e^{-ikr} \left\{ \left(ik - \frac{1}{r} \right) u - \frac{\partial u}{\partial r} \right\} \text{bounded}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (B)$$

are equivalent to one another and also to such intermediate conditions as, for instance,

$$u \rightarrow 0, \quad r \left(iku - \frac{\partial u}{\partial r}\right) \rightarrow 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (C)$$

$$ru \text{ bounded}, \quad r^2 \left(iku - \frac{\partial u}{\partial r}\right) \text{bounded}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (D)$$

Proof: It is evident, *a fortiori*, that (B) implies (A). But if (A) holds then, by the lemma,

$$\begin{aligned} r^3 e^{-ikr} \left\{ \left(ik - \frac{1}{r} \right) u - \frac{\partial u}{\partial r} \right\} &= -r^2 \frac{\partial}{\partial r} \left(u \frac{e^{ikr}}{r} \right) \\ &= -r^2 \frac{\partial}{\partial r} \sum_1^{\infty} a_n r^{1-n} \\ &= \sum_2^{\infty} (n-1) a_n r^{2-n}, \end{aligned}$$

and this is bounded, the series being uniformly convergent in the region $r > 3\rho_0$. Hence (A) implies (B), so that the two are equivalent. Similarly either of (C) and (D) implies (A) *a fortiori*, while (A) implies (C) and (D) on account of the lemma.

§ 4.

I prove now the uniqueness of the radiation problem as modified by Sommerfeld.

Theorem II. *There is no function u , not identically zero, which satisfies conditions (i)–(v) of § 1, except when $k=0$.*

Proof: Let $k=\alpha+i\beta$, and let Σ be as before a large sphere, centre X, radius R. We consider four cases separately.

(i) $\alpha \neq 0, \beta > 0$. We apply Green's theorem to u, \bar{u} , where \bar{u} denotes the complex conjugate of u , supposing that such a function u exists. Then

$$\iiint (u \nabla^2 \bar{u} - \bar{u} \nabla^2 u) dv = \iint_{S+\Sigma} \left(u \frac{\partial \bar{u}}{\partial n} - \bar{u} \frac{\partial u}{\partial n} \right) d\sigma = \iint_{\Sigma} \left(u \frac{\partial \bar{u}}{\partial R} - \bar{u} \frac{\partial u}{\partial R} \right) d\sigma,$$

since u and \bar{u} vanish on S. But on Σ we have, by the lemma,

$$u, \quad \frac{\partial u}{\partial R} = O(R^{-1}e^{-\beta R}),$$

and the same result holds for $\bar{u}, \partial \bar{u} / \partial R$. Making $R \rightarrow \infty$ we get

$$\iiint (u \nabla^2 \bar{u} - \bar{u} \nabla^2 u) dv = 0,$$

where the volume integral is extended over all space outside S. Hence, if \bar{k} denotes the complex conjugate of k ,

$$\iiint (k^2 - \bar{k}^2) u \bar{u} dv = 0$$

or

$$4i\alpha\beta \iiint |u|^2 dv = 0.$$

Since u is continuous this implies that u vanishes everywhere outside S, as was to be proved.

(ii) $\alpha=0, \beta>0$. We have, by Green's theorem in its first form,

$$\iiint (u \nabla^2 \bar{u} + |\text{grad } u|^2) dv = \iint_{S+\Sigma} u \frac{\partial \bar{u}}{\partial n} d\sigma.$$

As before, the surface integral over S vanishes, while on Σ

$$u, \frac{\partial \bar{u}}{\partial R} = O(R^{-1} e^{-\beta R}).$$

Hence, making $R \rightarrow \infty$, we get

$$\iiint (u \nabla^2 \bar{u} + |\text{grad } u|^2) dv = 0,$$

where the integral is extended over all space outside S. Since

$$\nabla^2 \bar{u} = \beta^2 \bar{u},$$

the required result follows as before.

(iii) $\beta < 0$. We have, by hypothesis, as $r \rightarrow \infty$,

$$u = O(r^{-1}), \quad r \left(iku - \frac{\partial u}{\partial r} \right) \rightarrow 0,$$

and hence $\partial u / \partial r = O(r^{-1})$. Hence

$$re^{-ikr} \left\{ \left(-ik - \frac{1}{r} \right) u - \frac{\partial u}{\partial r} \right\} \rightarrow 0,$$

as $r \rightarrow \infty$. That is to say, u satisfies the boundary conditions of Theorem I with $-k$ in place of k . This reduces the case $\beta < 0$ to the case $\beta > 0$, which has been dealt with already.

(iv) $\alpha \neq 0, \beta = 0$. It is in this case that we need the full force of the lemma. We have, as before,

$$\iiint (u \nabla^2 \bar{u} - \bar{u} \nabla^2 u) dv = \iint_{S+\Sigma} \left(u \frac{\partial \bar{u}}{\partial n} - \bar{u} \frac{\partial u}{\partial n} \right) d\sigma.$$

Here the volume- and the surface-integral over S both vanish so that

$$\iint_{\Sigma} \left(u \frac{\partial \bar{u}}{\partial R} - \bar{u} \frac{\partial u}{\partial R} \right) d\sigma = 0.$$

We have, by the lemma, on Σ ,

$$u = e^{ikR} \sum_1^{\infty} a_n R^{-n}, \quad \bar{u} = e^{-ikR} \sum_1^{\infty} \bar{a}_n R^{-n},$$

where the coefficients a_n depend only on $\text{grad } R$. Hence

$$\frac{\partial u}{\partial R} = iku - e^{ikR} \sum_1^{\infty} n a_n R^{-n-1},$$

with a corresponding result for \bar{u} , and so

$$\iint_{\Sigma} \left(-2ik \sum_1^{\infty} a_n R^{-n} \sum_1^{\infty} \bar{a}_n R^{-n} - \sum_1^{\infty} a_n R^{-n} \sum_1^{\infty} n \bar{a}_n R^{-n-1} \right. \\ \left. + \sum_1^{\infty} \bar{a}_n R^{-n} \sum_1^{\infty} n a_n R^{-n-1} \right) d\sigma = 0. \quad (4.1)$$

The leading term on the left-hand side is

$$- \frac{2ik}{R^2} \iint a_1 \bar{a}_1 d\sigma.$$

This must vanish, and hence

$$\iint |a_1|^2 d\sigma,$$

whence $a_1 \equiv 0$, since a_1 is continuous.

It may be proved by induction that the remaining coefficients vanish. Suppose that $a_m \equiv 0$ for $1 \leq m \leq n$. Then the leading term on the left-hand side of (4.1) will be

$$-2ikR^{-2n-2} \iint a_{n+1} \bar{a}_{n+1} d\sigma,$$

and this must vanish, whence $a_{n+1} \equiv 0$. Hence all the a_n must vanish, so that u vanishes identically.

This completes the proof of Theorem II. The case $k=0$ is that of potential theory, and the result does not hold.

§ 5.

While the above investigation appears to provide sufficient information on Sommerfeld's conditions for the purpose of diffraction problems, there are points of mathematical interest which remain unanswered. If, as is more natural from a mathematical point of view, we impose only one boundary condition at infinity we may get an eigen-value problem with solutions.

Thus for instance the condition

$$r^3 e^{-ikr} \left\{ \left(ik - \frac{1}{r} \right) u - \frac{\partial u}{\partial r} \right\} \text{ bounded}$$

may be written

$$\frac{\partial}{\partial r} \left(u \frac{e^{ikr}}{r} \right) = O \left(\frac{1}{r^2} \right),$$

and thus requires, roughly speaking, that u should behave for large r in the same way as a "fundamental solution" of the differential equation involved. As we have shown, the corresponding eigen-value problem will have no solutions for which the imaginary part of k is zero or positive.

However, if the surface S is a sphere every negative imaginary value of k will be an eigen-value. Write $k=i\beta$ where $\beta<0$, and let r be measured from the centre of the sphere S . Then the functions $r^{-1}e^{\beta r}$, $r^{-1}e^{-\beta r}$ both satisfy the differential equation and the boundary condition at infinity, and we can always choose a linear combination of these two functions which will vanish on S .

REFERENCES.

- (1) A. Sommerfeld, *Jahresbericht der D.M.V.* xxi. pp. 309-353 (1912)
- (2) A. Sommerfeld, 'Die Differential- und Integralgleichungen der Mechanik und Physik,' vol. II. Chapter 19, edited by P. Frank and R. von Mises, 2nd edition published by Mary S. Rosenberg, New York, in 1943.
- (3) B. B. Baker and E. T. Copson, 'Huygen's Principle,' Oxford (1939).

LXII. *On the use of Canonical Transformations for Collision Problems.*

By Professor W. HETTLER, F.R.S., and Dr. S. T. MA,
Dublin Institute for Advanced Studies*.

[Received March 30, 1949.]

SUMMARY.

A perturbation theory for collisions between free particles is developed by means of a canonical transformation with the following features:

- (i) The state of a free particle is redefined by inclusion of its virtual quanta.
- (ii) The self-energies of all states are included in the definition of the energies.
- (iii) The amplitude of the collision cross-section is given by an integral equation whose kernel is the transformed Hamiltonian and is given by an expansion in powers of the coupling parameter. A simplified derivation for the higher order corrections to the Compton effect is given and the analogous problem of meson scattering is discussed.

INTRODUCTION.

A GREAT deal of work has already been done on the higher order corrections to collision processes such as the Compton-effect or the scattering of a particle in a fixed potential. As was first realized by Bloch and Nordsieck (1937) (see also Pauli and Fierz 1938), it is essential, for a consistent

* Communicated by the Authors.

treatment of these higher order corrections, to re-define the states of a free particle in such a way that the virtual quanta which the particle is capable of emitting and absorbing are included in the definition of the state of the free particle. This re-definition is effected by a canonical transformation which, for a slowly moving particle interacting with the electromagnetic field, can be carried out exactly. For relativistic and mesonic problems an exact transformation has not been found yet in closed form, and one has to resort to an expansion in powers of the coupling parameter. The method will have to be applied for all problems whenever one is interested in higher orders. For eigenvalue problems the treatment is identical with the well-known perturbation theory. The case of the interaction of two particles has been considered by Pauli (1943) and Stueckelberg (1940). Although many authors have considered collision problems recently, it seems that a general theory has not been worked out yet (or at least has not been published yet), although in such a general theory hardly any new points or difficulties arise. Most authors confine themselves to an elimination of virtual quanta in first approximation only.

For collision problems, it had been realized by Wilson (1941), Gora (1943), Solokow (1941), Peng and one of us (Heitler 1941, Heitler and Peng 1942), that not all of the higher order corrections are due to virtual processes (in their work called "round-about transitions") but are due to the high degeneracy of the continuous spectrum (*cf.* also Pauli 1946). This so-called "damping effect" is there even in the first non-vanishing order (when the Hamiltonian, not the wave function, is expanded) and has no connection with the virtual quanta accompanying the free particles. The separation of the two types of effect is very natural and will appear to be so also in our present treatment. Although in this theory of damping use has usually been made only of the first order of the Hamiltonians (excluding the corrections due to round-about transitions), general formulæ for the higher corrections were given. These, however, were based on the idea of a "bare particle", not accompanied by virtual quanta, and would give rise to difficulties if actually used. The present treatment will give rise to a modification of these formulæ in higher orders and will free them of these difficulties.

For the treatment of the higher corrections (*i. e.* the virtual processes) the recent subtraction methods developed by Schwinger (1948) and Tomonaga have to be applied, if divergences are to be avoided (*cf.* also Lewis 1948 and Epstein 1948). They have been formulated in the most general way by Tomonaga and his pupils (Koba and Tomonaga 1947, 1948, Tati and Tomonaga (in the press)) *. While the complete neglect of all round about-transitions (as was done in the damping theory) is far too severe a subtraction, these authors have realized that all divergences are eliminated if only three invariant quantities are corrected (at least in

* We are very much indebted to Prof. Tomonaga for sending us manuscripts of this paper and of several others by Koba, Takeda, Fukuda, Miyamoto, Endo, Kinoshita (*Progr. Theor. Phys., Japan* (in the press)).

quantum electro-dynamics), namely: the self-mass, the self-charge of the particles and the self-mass of the photon. These corrections reduce to a re-normalization of these quantities. We shall embody the subtraction of the self-energies in our treatment in a very simple way, whereas the subtraction of vacuum-polarization effects would require a more detailed treatment of the Hamiltonian. The subtraction of an electromagnetic mass of the particles may be accepted as a sound—though preliminary—procedure, but the effects of vacuum-polarization (self-charge of the electron, self-mass of the photon) are more objectionable. They are connected with a deep-lying ambiguity of quantum-electrodynamics (*cf.*, for instance, Wentzel 1948, Ma (in the press)) *. This paper is concerned with a general, more formal, treatment of collision processes and these deeper questions will therefore not be entered upon. The formulæ given here will therefore not be entirely free from divergences.

A calculation based on lines similar to the present ones, for the special case of a scattering of an electron in a given potential, has been given some time ago by C. Morette and N. Hu (unpublished), but, to our knowledge these authors have not worked out a general theory †. We shall confine ourselves here to collisions between free particles, the case of an additional external field requires only a trivial generalization.

§ 1. TRANSFORMATION OF THE HAMILTONIAN.

We consider a representation in which the unperturbed Hamiltonian $H^{(0)}$, say, is diagonal with eigenvalues $E^{(0)}(\xi)$. $E^{(0)}(\xi)$ is the energy of free quanta and particles and depends in general on a continuously varying set of variables ξ (for instance the momenta of all particles). ξ shall be a complete set of variables thus specifying the unperturbed states of the system completely. Then:

$$\langle \xi' | H^{(0)} | \xi'' \rangle = E^{(0)}(\xi') \delta(\xi' - \xi''). \quad (1)$$

In order to simplify the notation and to indicate the analogy with discrete sets of variables, we write (1) in the form

$$H_{mn}^{(0)} = E_n^{(0)} \delta_{mn}, \quad (2)$$

with indices m, n in place of the variable ξ . A matrix product is to be understood as

$$\sum_k A_{mk} B_{kn} = \int d\xi \langle \xi' | A | \xi \rangle \langle \xi | B | \xi'' \rangle,$$

if m, n stand for ξ', ξ'' respectively.

Let the total Hamiltonian be

$$H = H^{(0)} + \Delta H, \quad (3)$$

where ΔH is the interaction.

* We are also indebted to Prof. Pauli for a comment on this point.

† We are very much indebted to Drs. Morette and Hu for their communication. (*Cf.* also Sebe 1948.)

We apply now to H a canonical transformation which will be specified by certain conditions later.

$$S^\dagger H S = \tilde{H}, \quad (4)$$

with the unitarity condition

$$S^\dagger S = 1, \quad (5)$$

so that

$$H S = S \tilde{H}. \quad (6)$$

The transformed Hamiltonian \tilde{H} can be separated into a diagonal part $E_{mn} \delta_{mn}$ and a non-diagonal part K_{mn} . Here the word diagonal means : diagonal with respect to *all* variables ; whereas K_{mn} may be diagonal with respect to some variables (as indeed we shall require below) but is non-diagonal with respect to at least one variable.

$$\tilde{H}_{mn} = E_{mn} \delta_{mn} + K_{mn}, \quad (7)$$

K_{mn} will be varying continuously with n , also in the neighbourhood of $n=m$; we may therefore except the point $n=m$ and put $K_{mn}=0$. Both E_m and K_{mn} will be functions of continuously varying variables m . For E we can write

$$E_m = E_m^0 + \Delta E_m. \quad (8)$$

ΔE_m will be the *self-energy* of the system in the state m . (6) can now be written in the form, with the help of (2), (3), (7), (8) :

$$E S - S E = S K + (\Delta E - \Delta H) S. \quad (9)$$

We depart here from the usual scheme of perturbation theory where the unperturbed energies $E^{(0)}$ are used as basic quantities through which everything is expressed. Instead we use the perturbed energies E , which include the self-energies of all states, because it is these energies which are the measured energies of free particles and quanta whereas the unperturbed energies $E^{(0)}$ have no measurable meaning *. In (9), and henceforth, only E will occur, and ΔE will be determined by a simple condition.

We expand ΔH , ΔE , K , S in a power series according to the powers of the coupling parameter involved in ΔH . In all quantum theories of fields ΔH can thus be expanded (with usually only the first two terms different from zero, sometimes even only the first) and then ΔE , K , S can also be expanded .

$$\Delta H = H^{(1)} + H^{(2)} + \dots, \quad (10 a)$$

$$\Delta E = E^{(1)} + E^{(2)} + \dots, \quad (10 b)$$

$$K = K^{(1)} + K^{(2)} + \dots, \quad (10 c)$$

$$S = 1 + S^{(1)} + S^{(2)} + \dots \quad (10 d)$$

* A similar perturbation theory has already been developed by Pirenne (1948), but Pirenne assumes all ΔE_m 's to be equal (independent of m) which is not sufficiently general.

Inserting (10) into (9) and equating terms of the same order, we obtain

$$ES^{(1)} - S^{(1)}E = K^{(1)} + E^{(1)} - H^{(1)}, \quad \dots \quad (11 a)$$

$$ES^{(2)} - S^{(2)}E = K^{(2)} + E^{(2)} - H^{(2)} + S^{(1)}K^{(1)} + (E^{(1)} - H^{(1)})S^{(1)}, \quad \dots \quad (11 b)$$

$$ES^{(3)} - S^{(3)}E = K^{(3)} + E^{(3)} - H^{(3)} + S^{(1)}K^{(2)} + S^{(2)}K^{(1)} + (E^{(2)} - H^{(2)})S^{(1)} + (E^{(1)} - H^{(1)})S^{(2)}, \quad \dots \quad (11 c)$$

$$ES^{(4)} - S^{(4)}E = K^{(4)} + E^{(4)} - H^{(4)} + S^{(1)}K^{(3)} + S^{(2)}K^{(2)} + S^{(3)}K^{(1)} + (E^{(3)} - H^{(3)})S^{(1)} + (E^{(2)} - H^{(2)})S^{(2)} + (E^{(1)} - H^{(1)})S^{(3)}, \quad \dots \quad (11 d)$$

and so on.

The unitarity condition requires that S can be expressed in the form

$$S = e^{iW}, \quad W = \text{hermitian}.$$

Thus, when W is expanded :

$$W = W^{(1)} + W^{(2)} + \dots, \quad \dots \quad (12 a)$$

$$S = 1 + iW - \frac{1}{2}W^2 - \frac{i}{6}W^3 + \dots, \quad \dots \quad (12 b)$$

$$S^{(4)} = iW^{(4)} + R^{(4)}, \quad \dots \quad (13 a)$$

$$R^{(1)} = 0, \quad R^{(2)} = -\frac{1}{2}(W^{(1)})^2, \quad R^{(3)} = -\frac{1}{2}(W^{(1)}W^{(2)} + W^{(2)}W^{(1)}) - \frac{i}{6}(W^{(1)})^3, \quad \dots \quad (13 b)$$

and so on.

So far S is, of course, quite arbitrary and K not determined. S has to be fixed so as to serve the purpose aimed at in the treatment of the particular problem considered. When we want to consider collisions between free particles and quanta, the total energy (always including the self-energies) of the free particles before and after collision are the same. We therefore impose the condition

$$K_{mn} = 0 \quad \text{if} \quad |E_m - E_n| > \epsilon, \quad \dots \quad (14)$$

where ϵ is a small quantity which will later tend to zero. In other words we make K_{mn} also nearly diagonal with respect to the total energy, but not with respect to other variables (momenta, directions and spins of individual particles, or indeed even the number of particles present). We call the energy region $|E_m - E_n| \leq \epsilon$ the energy shell. In the limit $\epsilon \rightarrow 0$, K_{mn} is diagonal with respect to the energy.

By (14), S is still not uniquely determined, for, obviously, the system is highly degenerate and we are still free to apply a transformation between the variables (other than the total energy) characterizing the system of

free particles. This we do not wish to do, because we are just interested in the transition probabilities between the states m , n , say, of the free particles. In other words we wish to maintain that the transformed wave function of the system

$$\tilde{\Phi}_m = \sum_n S_{mn} \Phi_n = \Phi_m + \sum_{\lambda > 1, n} i W_{mn}^{(\lambda)} \Phi_n + \sum_{\lambda > 1, n} R_{mn}^{(\lambda)} \Phi_n,$$

contains only contributions from *virtual* states n , with $|E_n - E_m| > \epsilon$, but no contribution from states of the same energy shell. This we would satisfy if we could make $S_{mn}^{(\lambda)} = 0$ for $\lambda \geq 1$ and $|E_m - E_n| < \epsilon$. This condition is, however, unfulfillable because it can easily be seen to contradict the unitarity condition for S . We may, however, in (15) also permit terms of the type $R_{mn}^{(\lambda)} \Phi_m$ or $W_{mm}^{(\lambda)} \Phi_m$ to occur, because these merely mean a renormalization of the original wave function Φ_m (which indeed is necessary if $\tilde{\Phi}_m$ is to be normalized). We shall therefore require

$$W_{mn}^{(\lambda)} = 0 \quad \text{for} \quad |E_m - E_n| < \epsilon, \quad . \quad . \quad . \quad (15)$$

and we shall show that then

$$R_{mn}^{(\lambda)} = R_m^{(\lambda)} \delta_{mn} + \text{a continuous function of } n \quad (\text{for } |E_m - E_n| < \epsilon) \quad . \quad . \quad . \quad (15')$$

The part of $R_{mn}^{(\lambda)}$ varying continuously with n at $n=m$, will not contribute to $\sum_n R_{mn}^{(\lambda)} \Phi_n$, when $\epsilon \rightarrow 0$. We might have permitted $W_{mm}^{(\lambda)} \neq 0$, but this is of no consequence because it merely means a phase transformation of Φ_m .

(15') holds only for collisions between free particles, and to show the validity of this equation it is necessary to make use of some properties of ΔH_{mn} . We use the fact that, for free particles, $\Delta H_{mn} = 0$ unless the total momentum is conserved in the transition $n \rightarrow m$. Now, by (13), $R_{mn}^{(\lambda)}$ is always expressed as a matrix product of at least two terms $W^{(\lambda_1)} W^{(\lambda_2)}$, or if W is expressed by $H^{(\lambda)}$, in the form $H^{(\lambda_1)} H^{(\lambda_2)}$. For instance,

$$R_{mn}^{(2)} \simeq \sum_k H_{mk}^{(1)} H_{kn}^{(1)}$$

(omitting certain energy denominators depending on k also). Now taking into account the conservation of momenta it is clear that: (i) when $m \neq n$, only a finite number of k 's can occur, (ii) when $m = n$, an infinite number of k 's occurs, involving an integration over an infinite range of values of some variables (momenta of virtual quanta, for instance) associated with k . Consequently, in the transformed wave function $\tilde{\Phi}$, the terms $R_{mn}^{(\lambda)} \Phi_n$ ($m \neq n$, $|E_m - E_n| < \epsilon$) will be negligible compared with $R_{mm}^{(\lambda)} \Phi_m$, and since the latter itself is of the order of Φ_m , $\sum_n R_{mn}^{(\lambda)} \Phi_n$ ($|E_m - E_n| < \epsilon$, $m \neq n$) will tend to zero, when $\epsilon \rightarrow 0$.

This consideration may serve as justification for fixing our canonical transformation by the condition (15). In the following we shall, however, not use (15') explicitly (we shall work with S directly rather than W and R), but it will appear again in the applications § 3 that $R_{mn}^{(2)}$ is negligible if $|E_m - E_n| < \epsilon$ and $m \neq n$.

(15), (15') give

$$S_{mn}^{(\lambda)} = R_{mn}^{(\lambda)}, \quad \text{for } |E_m - E_n| < \epsilon \quad \dots \quad (15'')$$

in all approximations. It is easily seen that S , K , and ΔE are now uniquely determined from (11), step by step: (11 a) first determines $S_{mn}^{(1)}$ for $|E_m - E_n| > \epsilon$ from $H^{(1)}$, and for $|E_m - E_n| < \epsilon$ the left-hand side vanishes (in the limit $\epsilon \rightarrow 0$) so $K^{(1)}$, $H^{(1)}$ are also determined from $H^{(1)}$ on the energy shell. The same procedure, applied to (11 b), ... determines all quantities in all approximations.

That S , determined in this way, is unitary requires perhaps a proof. This is very simple. We re-write (9) using W instead of S in the form

$$E + K = e^{-iW}(E - \Delta E + \Delta H)e^{iW},$$

and expand according to (12). This gives

$$K^{(1)} + E^{(1)} = H^{(1)} + i[EW^{(1)}],$$

$$K^{(2)} + E^{(2)} = H^{(2)} + i[EW^{(2)}] - \frac{1}{2}[[EW^{(1)}]W^{(1)}] + i[H^{(1)} - E^{(1)}, W^{(1)}], \quad \dots \quad (16)$$

and so on. We could determine W , K , ΔE also from (16) which is completely equivalent to (11). Now, a multiple commutator $[[A_1 A_2] A_3 \dots]$ is hermitian or antihermitian according as it contains an odd or even number of hermitian quantities and, vice versa, if $i[AB]$ and A are hermitian, B must be hermitian. Since $H^{(1)}$ is hermitian and outside the energy shell $K^{(1)} + E^{(1)} = 0$, $W^{(1)}$ is hermitian. On the energy shell $[EW^{(1)}]$ vanishes, therefore $K^{(1)} + E^{(1)}$ is hermitian, and the separation into diagonal and non-diagonal parts does not interfere with the hermiticity condition. Proceeding in the same way for $W^{(2)}$ we find that $W^{(2)}$, $K^{(2)}$, $E^{(2)}$ are hermitian and so on. W is therefore hermitian and S unitary.

§ 2. EXPLICIT EXPRESSIONS.

In applications to the quantum theory of fields it is sufficient to assume

$$\Delta H = H^{(1)} + H^{(2)}, \quad \dots \quad (17)$$

and furthermore $H^{(1)}$ has usually only matrix-elements for a pair of states of unequal energies. Hence

$$H_{mn}^{(1)} = 0 \quad \text{for } |E_m - E_n| < \epsilon. \quad \dots \quad (18)$$

We now derive the explicit expressions for S , K , ΔE up to the fourth order of K , ΔE . To indicate the distinctions of states on and outside the energy shell also in writing, we denote states with the same energy by A , B , ... whereas states with different energies are denoted by small letters :

$$|E_A - E_B| < \epsilon \quad |E_A - E_n| > \epsilon. \quad \dots \quad (19)$$

Furthermore, we use the notation

$$\omega_{mn} \equiv E_m - E_n. \quad \dots \quad (19')$$

(11 a) gives

$$E^{(1)} = K^{(1)} = 0, \quad S_{mn}^{(1)} = -\frac{H_{mn}^{(1)}}{\omega_{mn}}. \quad (20 a)$$

From (11 b), using (20 a) :

$$K_{AB}^{(2)} + E_{AB}^{(2)} = H_{AB}^{(2)} + \frac{H_{An}^{(1)}H_{nB}^{(1)}}{\omega_{Bn}}, \quad (20 b)$$

$$S_{mn}^{(2)} = -\frac{H_{mn}^{(2)}}{\omega_{mn}} + \frac{H_{mk}^{(1)}H_{kn}^{(1)}}{\omega_{mn}\omega_{kn}}. \quad (20 c)$$

A summation is to be understood over matrix-indices occurring twice. Since $S_{mn}^{(1)} = 0$ on the energy shell, terms for which any ω vanishes are to be omitted. If any summation occurs over a continuous range of states including those for which ω would vanish, $1/\omega$ is to be understood to mean the principal value of $1/\omega$. In (20 b), $E_{AB}^{(2)} = 0$ by definition if $A \neq B$ and $K_{AA}^{(2)} = 0$. $S_{AA}^{(2)}$ is still undetermined. We obtain it from (13 a, b) and the condition (15) imposed on W .

$$\begin{aligned} S_{AB}^{(2)} &= -\frac{1}{2}(W^{(1)})_{AB}^2 = +\frac{1}{2}(S^{(1)})_{AB}^2 \\ &= \frac{1}{2} \frac{H_{An}H_{nB}}{\omega_{An}\omega_{nB}} = R_{AB}^{(2)}. \end{aligned} \quad (20 d)$$

(20) are the familiar formulæ for the second order transitions, at least in appearance. They are, however, fundamentally different in that the energy denominators ω are the differences of the observable energies which include the self-energies and not the so-called unperturbed energies. This fact will lead to a different type of formulæ for the higher orders.

Proceeding in the same way for (11 c), (11 d) we obtain, making use of (18),

$$K_{AB}^{(3)} + E_{AB}^{(3)} = \frac{H_{An}^{(2)}H_{nB}^{(1)}}{\omega_{Bn}} + \frac{H_{An}^{(1)}H_{nB}^{(2)}}{\omega_{Bn}} + \frac{H_{An}^{(1)}H_{nk}^{(1)}H_{kB}^{(1)}}{\omega_{Bn}\omega_{Bk}}, \quad (21 a)$$

$$\begin{aligned} S_{mn}^{(3)} &= \frac{1}{\omega_{nm}} \left\{ \frac{H_{mk}^{(1)}K_{kn}^{(2)}}{\omega_{mk}} + E_n^{(2)} \frac{H_{mn}^{(1)}}{\omega_{mn}} + \frac{H_{mk}^{(2)}H_{kn}^{(1)}}{\omega_{nk}} + H_{mk}^{(1)}S_{kn}^{(2)} \right\} \\ &= \frac{H_{mn'}^{(1)}H_{n'n}^{(2)}}{\omega_{mn'}\omega_{nm}} + \frac{H_{mn'}^{(1)}H_{n'k}^{(1)}H_{kn}^{(1)}}{\omega_{mn'}\omega_{nk}\omega_{nm}} - \frac{H_{nn'}^{(2)}H_{mn}^{(1)}}{\omega_{nn'}^2} - \frac{H_{mk}^{(1)}H_{km}^{(1)}H_{mn}^{(1)}}{\omega_{mk}\omega_{km}\omega_{mn}} \\ &\quad + \frac{H_{mk}^{(2)}H_{kn}^{(1)}}{\omega_{nk}\omega_{nm}} + \frac{H_{mk}^{(1)}H_{kn}^{(2)}}{\omega_{nk}\omega_{nm}} + \frac{H_{mk}^{(1)}H_{kl}^{(1)}H_{ln}^{(1)}}{\omega_{kn}\omega_{ln}\omega_{nm}} + \frac{1}{2} \frac{H_{mn'}^{(1)}H_{n'k}^{(1)}H_{kn}^{(1)}}{\omega_{n'k}\omega_{kn}\omega_{nm}}, \end{aligned} \quad (21 b)$$

with $n' \neq n$, $|\omega_{nn'}| < \epsilon$, $|\omega_{nn''}| < \epsilon$.

Here states with energy equal to E_n (within the shell ϵ) are denoted by n' , n'' . m is, of course, no summation index. Otherwise the rules of summation apply. Whilst $n' = n$ is excluded, n'' in the last term (which arises from $S_{nn}^{(2)}$ with $|\omega_{nn''}| < \epsilon$) may, but need not, be equal to n .

$S_{AB}^{(3)}$ could be calculated from (12), (13) but will not be needed except in orders higher than the fourth. Finally, we get from (11 d)

$$(K^{(4)} + E^{(4)})_{AB} = \frac{H_{Ak}^{(1)} K_{kB}^{(3)}}{\omega_{Ak}} - S_{AC}^{(2)} K_{CB}^{(2)} + H_{Ak}^{(2)} S_{kB}^{(2)} \\ + H_{AC}^{(2)} S_{CB}^{(2)} - E_A^{(2)} S_{AB}^{(2)} + H_{Ak}^{(1)} S_{kB}^{(3)}.$$

In the term $H^{(2)} S^{(2)}$ we have distinguished between states k, C with $|\omega_{Ak}| > \epsilon$, $|\omega_{AC}| < \epsilon$. The first term vanishes because either $H_{Ak}^{(1)} = 0$ (when $|\omega_{Ak}| < \epsilon$, (18)) or $K_{kB}^{(3)} = 0$ (when $|\omega_{kB}| > \epsilon$). In the last term only states $|\omega_{Ak}| > \epsilon$ occur. Inserting (20 b, c, d) and (21 b) we obtain

$$K_{AB}^{(4)} + E_{AB}^{(4)} = \left(H_{An}^{(2)} + \frac{H_{Ak}^{(1)} H_{kn}^{(1)}}{\omega_{Ak}} \right) \frac{1}{\omega_{Bn}} \left(H_{nB}^{(2)} + \frac{H_{nl}^{(1)} H_{lB}^{(1)}}{\omega_{Bl}} \right) \\ + \frac{H_{Ak}^{(1)} H_{kl}^{(2)} H_{lB}^{(1)}}{\omega_{Bk} \omega_{Bl}} - \frac{1}{2} \frac{H_{Ak}^{(1)} H_{kC}^{(1)}}{\omega_{Ak}^2} \left(H_{CB}^{(2)} + \frac{H_{Cn}^{(1)} H_{nB}^{(1)}}{\omega_{Bn}} \right)_{C \neq B} \\ - \frac{1}{2} \left(\frac{H_{Ak}^{(1)} H_{kC}^{(1)}}{\omega_{Ak}} + H_{AC}^{(2)} \right)_{C \neq A} \frac{H_{Cn}^{(1)} H_{nB}^{(1)}}{\omega_{Bn}^2} \\ - \frac{H_{Ak}^{(1)} H_{kB}^{(1)}}{\omega_{Ak}^2} \left(H_{kk}^{(2)} + \frac{H_{kn}^{(1)} H_{nk}^{(1)}}{\omega_{kn}} \right).$$

Here we have used $\omega_{Ak} = \omega_{Bk}$, etc., on account of $|E_B - E_A| < \epsilon$. A number of terms then cancel or can be comprised. In the third term $C=B$ has to be excluded but $C=A$ is included. Similarly, in the fourth term $C=A$ is excluded but $C=B$ included. The last term contains the self-energy of the state k . This also occurs in the first and second terms, because here the summation includes also $k=l$. The corresponding terms can then be comprised, using

$$\frac{1}{\omega_{Ak}^2 \omega_{Bn}} - \frac{1}{\omega_{Ak}^2 \omega_{kn}} = - \frac{1}{\omega_{Ak} \omega_{Bn} \omega_{kn}}.$$

The final result is :

$$K_{AB}^{(4)} + E_{AB}^{(4)} = \left(H_{An}^{(2)} + \frac{H_{Ak}^{(1)} H_{kn}^{(1)}}{\omega_{Ak}} \right) \frac{1}{\omega_{Bn}} \left(H_{nB}^{(2)} + \frac{H_{nl}^{(1)} H_{lB}^{(1)}}{\omega_{Bl}} \right)_{l \neq k} \\ + \frac{H_{Ak}^{(1)} H_{kl}^{(2)} H_{lB}^{(1)}}{\omega_{Bk} \omega_{Bl}} \Big|_{l \neq k} - \frac{H_{Ak}^{(1)} H_{kn}^{(1)} H_{nk}^{(1)} H_{kB}^{(1)}}{\omega_{Ak} \omega_{Bn} \omega_{kn}} \\ - \frac{1}{2} \left(\frac{H_{Ak}^{(1)} H_{kC}^{(1)}}{\omega_{Ak}} + H_{AC}^{(2)} \right)_{C \neq A} \frac{H_{Cn}^{(1)} H_{nB}^{(1)}}{\omega_{Bn}^2} \\ - \frac{1}{2} \frac{H_{Ak}^{(1)} H_{kB}^{(1)}}{\omega_{Ak}^2} \left(H_{CB}^{(2)} + \frac{H_{Cn}^{(1)} H_{nB}^{(1)}}{\omega_{Bn}} \right)_{C \neq B} \quad \dots \quad (22)$$

In (22) no term occurs anywhere that could be reduced to the self-energy of any state. The self-energies are all eliminated from the expression for $K_{AB}^{(4)}$. Of course, the diagonal element $A=B$ of (22) represents itself the

fourth order self-energy of the state A. This is somewhat different from the expression usually found in the literature, and the reason is that the second order self-energy contains the true energies and not the unperturbed energies in the denominator.

It will appear that the last two terms, which are proportional to $R_{CB}^{(2)}$ and $R_{AC}^{(2)}$ respectively, will reduce to contributions $\sim R_{EB}^{(2)}$ and $R_{AA}^{(2)}$, in agreement with (15').

§ 3. THE COLLISION INTEGRAL EQUATION.

The wave equation for stationary processes is generally given by

$$E\Phi = H\Phi,$$

where E is the energy of the system. After carrying out the canonical transformation of § 1 this becomes

$$E\tilde{\Phi} = \tilde{H}\tilde{\Phi},$$

where $\tilde{\Phi} = S^{-1}\Phi$ is the transformed wave function. With the help of (7) this becomes

$$(E - E_n)\tilde{\Phi}_n = K_{nm}\tilde{\Phi}_m. \quad (23)$$

We shall henceforth take the view that it is the transformed wave function $\tilde{\Phi}_m$ which describes the true state of a system of free particles. Each particle is then considered to be accompanied by all its virtual quanta which it is capable of emitting, a fact which has first been fully recognized by Bloch and Nordsieck (1937). Accordingly, a collision between free particles will be described as a transition between two $\tilde{\Phi}$'s and not two Φ 's. Written out fully, (23) reads:

$$(E - E(\xi))\langle \xi | \tilde{\Phi} | \xi_0 \rangle = \langle \xi | K | \xi' \rangle \langle \xi' | \Phi | \xi_0 \rangle, \quad (24)$$

where we have added a symbol ξ_0 on the right of Φ to indicate a particular initial state ξ_0 . If we consider transitions between free particles only, energy will be strictly conserved and we may put *

$$E = E(\xi_0). \quad (25)$$

If ξ_0 is the initial state and all other states are outgoing waves $\tilde{\Phi}$ takes the well-known form

$$\langle \xi | \tilde{\Phi} | \xi_0 \rangle = \delta(\xi - \xi_0) + \langle \xi | U | \xi_0 \rangle \left\{ \frac{P}{E(\xi_0) - E(\xi)} - i\pi\delta(E(\xi_0) - E(\xi)) \right\}. \quad (26)$$

Here U is an unknown matrix which however can be supposed to be non-singular at $E(\xi_0) = E(\xi)$. P denotes the principal value. It will be convenient to separate the variables ξ into the total energy and all other

* The case of collisions involving bound states where E is an independent variable which must not be identified with any of the states E_m of the system has recently been considered by us (Heitler and Ma 1949).

variables x , say, such as the number of particles present, their momenta, distribution, spins, etc. Inserting (26) into (24) we obtain then

$$\begin{aligned} \langle Ex | U | E_0 x_0 \rangle &= \langle Ex | K | E_0 x_0 \rangle \\ &+ \int_{E-\epsilon}^{E+\epsilon} dE' \int dx' \langle Ex | K | E' x' \rangle \langle E' x' | U | E_0 x_0 \rangle \cdot \frac{P}{E_0 - E'} \\ &- i\pi \int dx' \langle Ex | K | E_0 x' \rangle \langle E_0 x' | U | E_0 x_0 \rangle. \quad \dots \quad (27) \end{aligned}$$

The limits $E-\epsilon$, $E+\epsilon$ in the second term arise from the fact that K is different from zero only in this interval.

U , whose square is essentially the collision cross-section, will be required only for $E=E_0$. Since both K and U are non-singular at $E_0=E'$, the integral over E' becomes for $\epsilon \rightarrow 0$

$$\int_{E_0-\epsilon}^{E_0+\epsilon} \frac{P}{E_0 - E'} dE' = 0. \quad \dots \quad (28)$$

Thus the second term of (27) disappears. Returning to the notation of § 2 we obtain for U the integral equation

$$U_{AO} = K_{AO} - i\pi K_{AB} U_{BO}. \quad \dots \quad (29)$$

This is the integral equation previously derived for collision processes by Wilson, Peng and Heitler and others (*loc. cit.*). In their applications use has been made only of the lowest order of K —in the simplest cases $K^{(2)}$, for more complicated processes the first non-vanishing term of the series $\Sigma K^{(k)}$ —owing to the divergence of the higher terms. However, the expression for K obtained here differs in the higher orders from those previously derived. The expression previously given was, for the case $H^{(2)}=0$,

$$K_{AB}^{(2)} = \frac{H_{An} H_{nB}}{\omega_{Bn}}, \quad K_{AB}^{(4)} = \sum_{n, m, k} \frac{H_{An} H_{nm} H_{mk} H_{kB}}{\omega_{Bk} \omega_{Bm} \omega_{Bn}}, \quad \dots \quad (30)$$

which is quite different from (22). The difference is due to two reasons: (30) is derived with the understanding that the initial state is the unperturbed state B of bare particles not accompanied by their virtual quanta, and furthermore, the ω_{Bn} in $K^{(2)}$ are the energy-differences of the unperturbed energies not corrected by the self-energies. The use of (30) for the higher corrections of K would be impossible on account of singularities occurring, when, for instance, $k=n$. The present treatment, although not completely freed from all divergences, is preferable both for mathematical and physical reasons.

When considering, for instance, the simplest collision, that between two free particles, $K_{AB}^{(2)} \neq 0$. The replacement of U_{AO} by $K_{AO}^{(2)}$ is the Born-Dirac approximation, commonly used for the calculation of the Compton effect, etc. It appears that the exact treatment leads to two different corrections: (i) we have to include $K_{AB}^{(4)}$, etc., and (ii) we have to include the "damping term"— $i\pi K_{AB} U_{BO}$. *A priori* it is impossible to say which of the two corrections is more important, and this depends also on the particular problem involved. The first correction $K_{AB}^{(4)}$ has only recently

become accessible to treatment by means of the recent subtraction methods. In particular, for the Compton-effect, Jost and Corinaldesi (Corinaldesi and Jost 1948) have evaluated $K_{AB}^{(4)}$ explicitly (for the non-relativistic region). The "damping effect" had previously been considered by S. Power (1945), as far as the term $K^{(2)}U$ is concerned, which is always finite, and has been found to be negligible for all energies (in particular also in the relativistic region). For meson processes the situation is quite different and is probably the reverse.

From the solution of (21), the cross-section for the process $O \rightarrow A$ follows in the well-known way and is

$$\Phi_{AO} = \frac{2\pi}{v} |U_{AO}|^2 \rho_A,$$

where ρ_A is the density function of the states A , v the relative velocity of the colliding particles in the state O .

It may be of some interest, as an illustration of the present formalism, to re-derive the formula of Jost and Corinaldesi in a very simple way, as the original derivation based on a relativistic treatment, is extremely complicated. This will be done in the following section.

§ 4. HIGHER CORRECTIONS TO THE COMPTON-EFFECT.

We consider now the corrections to the scattering formula of a photon by a free electron. The relativistic calculations are extremely involved and we confine ourselves to the non-relativistic case $\hbar v \ll mc^2$. The first non-vanishing term is $K^{(2)}$, which gives the usual Klein-Nishina formula, or, in non-relativistic approximation, the Thomson formula (with some correction of the order $\hbar v/mc^2$ which even in the non-relativistic treatment can be obtained correctly). Non-relativistically the Hamiltonian of the interaction is

$$H^{(1)} = -\frac{e(\mathbf{p}\mathbf{A})}{m_0}, \quad H^{(2)} = \frac{e^2 \mathbf{A}^2}{2m_0}. \quad (23)$$

We assume the electron to be initially at rest so that, if the initial state is denoted by 0,

$$H_{n0}^{(1)} = 0, \quad (23')$$

for all states n . The primary (secondary) photon is denoted by k_0 (k), whereas virtual photons are denoted by k' , etc. We choose units $\hbar=c=1$. It will be necessary to take always the recoil energy $p^2/2m$ of the electron into account, but we do this also in non-relativistic approximation only. (23) applies to particles with spin 0 only, for particles with spin $\frac{1}{2}$ an additional term $-\mu_0(\sigma\mathbf{H})$ would have to be added to H in the approximation considered (μ_0 =Bohr magneton).

It should be noticed that the electron mass occurs in (23) explicitly. m_0 is, by definition, the mechanical mass and not the true mass. Although we have replaced the energies throughout by their true values, we have not made any provision for a mass correction occurring explicitly in H . To this point we shall return below.

If m, n are states (not necessarily on the same energy shell) with one photon present and the electron with such momenta as are in agreement with the conservation of momentum

$$H_{mn}^{(2)} = \frac{2\pi e^2}{m_0} (\mathbf{e}_m \mathbf{e}_n) \frac{1}{\sqrt{(k_m k_n)}}, \quad (24 a)$$

where \mathbf{e} is the polarization vector of the photon. We obtain then from (20 b) and (23')

$$K_{AO}^{(2)} = H_{AO}^{(2)} = \frac{2\pi e^2}{m_0} (\mathbf{e}_0 \mathbf{e}) \frac{1}{\sqrt{(k_0 k)}}, \quad \mathbf{k}_0 = \mathbf{p} + \mathbf{k}. \quad (24 b)$$

It is not necessary to neglect \mathbf{p} against \mathbf{k}_0 , up to the first order in p/k_0 . The matrix-elements $H_{mn}^{(1)}$ for emission or absorption of one photon \mathbf{k}' are

$$H_{mn}^{(1)} = -e \sqrt{\left(\frac{2\pi}{k'}\right)} (\mathbf{p} \mathbf{e}'). \quad (24 c)$$

The formula (22) for $K_{AO}^{(4)}$ reduces to (on account of (23')):

$$K_{AO}^{(4)} = \frac{H_{An}^{(2)} H_{nO}^{(2)}}{\omega_{On}} + \frac{H_{Ak}^{(1)} H_{kn}^{(1)} H_{nO}^{(2)}}{\omega_{On} \omega_{Ok}} - \frac{1}{2} \frac{H_{Ak}^{(1)} H_{kC}^{(1)} H_{CO}^{(2)}}{\omega_{Ok}^2} \Big|_{C \neq O}. \quad (25)$$

If A is the scattered state, the states n occurring in the first and second terms of (25) can be

$$n : \mathbf{k}', \mathbf{p}' = \mathbf{k}_0 - \mathbf{k}' \quad \text{or} \quad n : \mathbf{k}_0, \mathbf{k}, \mathbf{k}', \mathbf{p}' = -\mathbf{k}' - \mathbf{k},$$

where \mathbf{k}' is a photon with arbitrary momentum (and energy). Then :

$$\frac{H_{An}^{(2)} H_{nO}^{(2)}}{\omega_{On}} = \frac{1}{2\pi} \frac{e^4}{m_0^2} \frac{\Sigma (\mathbf{e}_0 \mathbf{e}') (\mathbf{e}' \mathbf{e})}{\sqrt{(k_0 k)}} \int k' dk' d\Omega_{k'} \cdot \left[\frac{1}{k_0 - k' - \frac{(\mathbf{k}_0 - \mathbf{k}')^2}{2m}} - \frac{1}{k + k' + \frac{(\mathbf{k} + \mathbf{k}')^2}{2m}} \right].$$

$d\Omega_{k'}$ is the element of the solid angle for the direction of \mathbf{k}' . We wish to evaluate this up to terms $1/m^3 \log m$. No higher accuracy is justified. It is then easily seen that $(\mathbf{k}_0 - \mathbf{k}')^2/2m$ can be replaced by $k'^2/2m$, and similarly for $(\mathbf{k} + \mathbf{k}')^2/2m$. The evaluation is then very simple and yields (in the bracket we make no distinction between m_0 and m)

$$\frac{H_{An}^{(2)} H_{nO}^{(2)}}{\omega_{On}} = - \frac{e^4}{m_0} (\mathbf{e}_0 \mathbf{e}) \frac{1}{\sqrt{k_0 k}} \frac{4}{3} \left\{ 4 \log \frac{K'}{2m} \Big|_{K' \rightarrow \infty} + \frac{(3 + \cos \theta) k_0^2}{m^2} \log \frac{k_0}{2m} \right\}. \quad (26 a)$$

We obtain a logarithmically divergent term, which will be discussed below, and a finite term. Terms $1/m^3$ without logarithms have not been included. θ is the scattering angle.

In much the same way, the second term of (25) can be—very simply—evaluated and yields

$$\frac{H_{Ak}^{(1)} H_{kn}^{(1)} H_{n0}^{(2)}}{\omega_{0n} \omega_{0k}} = \frac{e^4}{m^3} \frac{8}{3} \frac{(\mathbf{e}_0 \mathbf{k})(\mathbf{e} \mathbf{k}_0)}{k_0} \log \frac{k_0}{2m}. \quad (26 b)$$

Finally we consider the last term of (25). If $C \neq A$, the transition $A \rightarrow C$ can only occur by means of a finite number of intermediate states k with photons of fixed momentum. If the summation over C is limited to an energy band of vanishing width this term vanishes. This is in agreement with (15'), for the first factor is nothing but $R_{AC}^{(2)}$. This is, however, not so when $C=A$. Then the states k can be virtual photons with arbitrary momentum. If evaluated in the same way as before, it turns out that

$$\Sigma \frac{H_{Ak}^{(1)} H_{kA}^{(1)}}{\omega_{0k}^2}$$

diverges logarithmically for low values of the virtual k' . This is closely connected with the infra-red problem. We call the lower limit of integration for the moment k_m and find

$$-\frac{1}{2} \frac{H_{Ak}^{(1)} H_{kA}^{(1)}}{\omega_{0k}^2} H_{A0}^{(2)} = \frac{4}{3} \frac{e^4}{m^3} (\mathbf{e}_0 \mathbf{e}) (1 - \cos \theta) k_0 \log \frac{k_m}{2m}. \quad (26 c)$$

Also here we have neglected terms free from logarithms.

The diverging term in (26 a) appears to be exactly proportional to $K_{A0}^{(2)}$ (24 b) multiplied by a constant which neither depends on the initial nor the final state. The constant, though infinite, is purely numerical. This term will therefore be merely a numerical correction to the constants occurring already in $H_{A0}^{(2)}$. In agreement with the modern subtraction methods we may feel free to regard it as a correction to the mechanical mass m_0 which occurs explicitly in $K_{A0}^{(2)}$, and which so far has not been corrected. Naturally, such a procedure has to be justified by a thorough investigation of the mass (and charge) corrections which, of course, is only possible in a relativistic treatment*. We cannot expect from our purely non-relativistic treatment to obtain the correct mass and charge corrections. A relativistic justification for this procedure has, however, been given by Jost and Corinaldesi and, more completely by Tomonaga and his co-workers (*loc. cit.*). We may therefore omit the term in question and write m for m_0 .

If we neglect the damping term of (29) the cross-section is given by

$$\Phi_{A0} = \frac{1}{4\pi^2} [K_{A0}^{(2)} + K_{A0}^{(4)}]^2 \frac{k^3}{k_0} d\Omega.$$

After summing over \mathbf{e} and averaging over \mathbf{e}_0 we get, up to terms $\sim e^6$:

$$\Phi_{A0} = \Phi_4 + \Phi_6, \quad (27 a)$$

* The situation is actually slightly different in a relativistic treatment, because the mass m_0 does not occur there explicitly in the interaction ΔH , which is $-e(\alpha A)$ in Dirac's equation.

$$\Phi_4 = \left(\frac{e^2}{m}\right)^2 \frac{1 + \cos^2 \theta}{2} d\Omega \cdot \frac{k^2}{k_0^2}, \quad (27 b)$$

$$\Phi_6 = \frac{2}{3} \frac{d\Omega}{\pi} \frac{e^6}{m^2} \frac{k_0^2}{m^2} \left\{ (1 + \cos^2 \theta)(1 - \cos \theta) \log \frac{k_m}{2m} - (3 + 3 \cos \theta + 3 \cos^2 \theta - \cos^3 \theta) \log \frac{k_0}{2m} \right\}. \quad (27 c)$$

In Φ_4 it is consistent to keep the difference of k and k_0 up to terms k_0/m .

Finally, we have to discuss the logarithmic divergence due to $k_m \rightarrow 0$. This is completely compensated by the double Compton-effect, where in addition to the scattered quantum, a very soft quantum is produced. As Jost (1947) has already shown, the cross-section for this process is exactly equal to $-\Phi'_6$, where Φ'_6 is the part of Φ_6 proportional to $\log k_m$. This is also very easily checked by direct calculation. We may therefore identify k_m with a finite and somewhat arbitrary lower limit, namely, the limit at which we choose to consider the scattering with and without emission of a second extremely soft quantum as two different processes. That this limit is arbitrary lies in the nature of the problem. The formula (27 c) is completely identical with the formula of Jost and Corinaldesi derived through very complicated relativistic calculations*.

We would finally have to discuss the influence of the damping term in (29). In first approximation this can be replaced by $K_{AB}^{(2)} K_{BO}^{(2)}$ and has in this way been evaluated by S. Power (1945). This gives clearly a correction e^4 to K . However, owing to the factor i , and because $K^{(2)}$ is real, the e^6 correction to the cross-section cancels and only a correction $\sim e^8$ arises, which is of smaller order than the effects considered above. This is also true in the relativistic region, also here the correction due to the damping term is $\sim e^8$ and decreases in the same way as the Klein-Nishina formula and is therefore completely negligible. In this region we have so far no knowledge about the magnitude of the effects arising from $K^{(4)}$. A comparison of Φ_6 and Φ_4 shows that Φ_6 is smaller by a factor of the order $(k_0^2/137m^2) \log(m/k_0)$, or $\log(m/k_m)$. This is for all energies $k_0 < m$ also quite negligible.

In a very similar way we could also calculate the correction to the scattering of a meson by a nucleon. This we have actually done for longitudinal vector mesons and obtained: (i) a logarithmically divergent term very similar to the first term of (26 a), which is also proportional to $K^{(2)}$, and which is—presumably—to be cancelled by a renormalization of the constants occurring in $K^{(2)}$; (ii) finite contributions $K^{(4)}$. The latter differ from $K^{(2)}$ by a factor of the order $g^2 k M / \mu^2$ (k =meson momentum, μ =meson mass, M =nucleon mass, g^2 =coupling constant) when $k \ll M$. For k approaching M , the correction increases far less rapidly than k/μ , and eventually decreases again. For conventional values of g^2 it appears that

* Jost and Corinaldesi also gave the terms of the order K_0^2/m^2 free from logarithms, the present treatment would give slightly different numerical factors in these terms.

$K^{(4)}$ is negligible for very low values of k , but approaches the order of $K^{(2)}$ when $k \sim \frac{1}{2}M$, say. It would seem, if this result is reliable, that the expansion $K = K^{(2)} + K^{(4)} + \dots$ is not quickly converging for larger values of k . Here, however, as was shown previously (Heitler and Peng 1942), the damping term is much larger and preponderant.

We do not wish to discuss this problem yet in detail because, evidently, a relativistic treatment is required to justify the proper renormalization of the constants (and hence the suppression of the divergent term). A relativistic investigation by Corinaldesi is in progress.

REFERENCES.

- BLOCH and NORDSIECK, 1937, *Phys. Rev.*, **52**, 54.
 CORINALDESI and JOST, 1948, *Helv. Phys. Acta*, **21**, 183.
 EPSTEIN, 1948, *Phys. Rev.*, **73**, 177.
 GORA, 1943, *Z. Phys.*, **120**, 121.
 HEITLER, 1941, *Proc. Camb. Phil. Soc.*, **38**, 291.
 HEITLER and PENG, 1942, *Proc. Camb. Phil. Soc.*, **38**, 296.
 HEITLER and MA, 1949, *Proc. Roy. Irish Acad.*, **52**, 109.
 JOST, 1947, *Phys. Rev.*, **72**, 173.
 Koba and TOMONAGA, 1947, *Progr. Theor. Phys.*, Japan, **2**, 128; 1948, *ibid.*, **3**, 290.
 LEWIS, 1948, *Phys. Rev.*, **73**, 173.
 MA, 1949, *Phys. Rev.* (in the press).
 PAULI, 1943, *Phys. Rev.*, **64**, 332; 1946, *Meson Theory of Nuclear Forces* (New York: Interscience Publishers) p. 41ff.
 PAULI and FIERZ, 1938, *Nuovo Cim.*, **15**, 267.
 PIRENNE, 1948, *Helv. Phys. Acta*, **21**, 226.
 POWER, 1945, *Proc. Roy. Irish Acad.*, **50**, 139.
 SCHWINGER, 1948, *Phys. Rev.*, **73**, 415; *ibid.*, **74**, 1439.
 SEBE, 1948, *Progr. Theor. Phys.*, Japan, **3**, 304.
 SOKOLOV, 1941, *J. Phys.*, U.S.S.R., **5**, 231.
 STUECKELBERG, 1940, *Helv. Phys. Acta*, **13**, 347.
 TATI and TOMONAGA, 1949, *Progr. Theor. Phys.*, Japan (in the press).
 WENTZEL, 1948, *Phys. Rev.*, **74**, 1070.
 WILSON, 1941, *Proc. Camb. Phil. Soc.*, **37**, 291.

CORRIGENDUM

In Prof. Ewen M'Ewen's paper in the 'Philosophical Magazine,' Ser. 7, Vol. 40, No. 303, April (1949), the following correction should be made:—

Page 456, last line of equation (9),

$$\text{for } \mu\phi - z \frac{\partial\psi}{\partial z} - \mu z \frac{\partial\psi}{\partial z} \quad \text{read} \quad \mu\phi - z \frac{\partial\psi}{\partial y} - \mu z \frac{\partial\psi}{\partial z}.$$

LXIII. *Lattice Defects in Silver Halide Crystals.*

By J. W. MITCHELL, D.Phil.,

H. H. Wills Physical Laboratory, University of Bristol*.

[Received May 9, 1949.]

WE owe most of our experimental knowledge of electrical conductivity of silver halides and other ionic solids to the work of Tubandt and his collaborators (reviewed by Tubandt 1932). Tubandt's work seemed to establish that in silver halides the transport number for silver was unity, and on the basis of this result it has been widely supposed that the Frenkel mechanism (interstitial silver ions and vacant cation sites) is the correct one for these materials. On the other hand, the work of Kolthoff and O'Brien (1939) and of Langer (1943) on the exchange of radioactive ions with silver bromide microcrystals in an electrolyte or non-aqueous medium shows that the diffusion coefficients for silver and bromide ions are similar. This is not consistent with the existence of Frenkel defects nor indeed with Tubandt's results on any model which admits a similar mechanism for diffusion and conduction. The purpose of this note is then :

(a) To re-examine Tubandt's results for silver chloride and silver bromide and to suggest that they are at least consistent with the Schottky mechanism (conduction through mobile cation and anion vacant sites).

(b) To revise one feature of the model suggested in our previous paper (Mitchell 1949) for the defects responsible for the optical sensitivity of silver halides in the presence of silver sulphide.

Let us consider an experiment in which a slab of silver halide at a temperature between 200 and 400° C. is placed between silver electrodes across which a potential difference is applied. In the dark, an ionic current will flow and silver will separate at the cathode. We will now discuss the formation of this silver. If conduction is by the Frenkel mechanism, interstitial ions will approach the cathode. They are unlikely to produce silver threads within the silver halide for two reasons : (a) this could only occur through the splitting or plastic deformation of the surrounding halide ; and (b) the energy level of an electron trapped in the field of an interstitial ion is very high (Simpson 1949). Thus ions approaching the cathode would not receive electrons from the metal and therefore any "metallic" binding force tending to make the ions cohere to form internal silver would be absent. Massive silver should therefore separate at the cathode. The formation of vacant cation sites at the interface between the cathode and the silver halide should also produce massive silver.

* Communicated by the Author.

Now, in Tubandt's experiments with pure silver halides (Tubandt *et al.* 1914, 1921 a, 1921 b, 1932), metallic silver separated at the cathode as a reasonably coherent mass only with α silver iodide which has a random cation lattice (Strock 1934). The silver ions in α silver iodide should have properties similar to the interstitial ions of Frenkel defects and therefore, for the reasons given above, the formation of internal threads of silver would, in this case, not be expected.

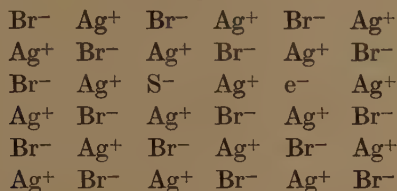
Internal threads are, however, formed very rapidly in silver chloride and silver bromide, and in the other modifications of silver iodide and this can readily be explained in terms of the Schottky mechanism. In this case, vacant anion sites approach the cathode. These (a) remove the halogen and (b) have low electronic levels (forming F centres) and so can receive electrons from the metal. Thus the formation of threads is definite evidence in favour of the Schottky mechanism.

We have then to ask why, in his experiments, Tubandt measured a transport number of unity for silver when a slab of α silver iodide was interposed between the cathode and the two slabs of silver chloride or silver bromide (Tubandt 1921 a). A possible explanation is that the current due to vacant anion sites is balanced by a flow of double vacancies in the reverse direction. This can only occur if their jump frequency is greater than that of a single vacancy by a factor greater than eV/kT , where V is the voltage across the slab. However, according to the calculations of Dienes (1948, see also Seitz 1946), this may well be the case. It will then be necessary to explain why this process is preferred to the stripping away of layers of halide. Not enough is known at present about the complicated conditions at the interface between the α silver iodide and the other silver halide to make such an attempt profitable. These arguments do not exclude the possible coexistence of interstitial silver ions and vacant anion lattice sites. We regard this, however, as unlikely and conclude that Schottky defects are predominant in the silver halides, apart from α silver iodide, and that both vacant cation and vacant anion sites are mobile.

Under these circumstances, an isolated F centre in a silver chloride or silver bromide crystal would also be mobile by the mechanisms discussed in our previous paper and such mobility would soon lead to aggregation. As both the quenched crystals of silver halides containing traces of silver sulphide, and the sensitive grains of high speed negative photographic emulsions made by sulphur sensitizing are stable at room temperature and do not form F centre aggregates, we must now assume that the F centres in these systems are present initially as immobile association complexes with singly charged sulphide ions (fig. 1). In this state, at low temperatures, their electrons will not be transferred to approaching vacant anion lattice sites by tunnel effect as occurs for isolated F centres. F centre pairs would have the same property. Under illumination, electrons will be released from these immobile F centres which in the case of the silver halide grains in photographic emulsions are likely to be

near the surface. The electrons will be captured by either previously formed F centre aggregates or else by vacant anion lattice sites thus producing mobile F centres which will themselves then form aggregates. At temperatures where vacant anion sites are mobile isolated F centres will have only a limited lifetime in dry silver halide crystals. This is

Fig. 1.



in accord with the experimental observations of Hilsch and Pohl (1930), who were unable to observe absorption due to isolated F centres. The conclusion will not of necessity apply to F centres in the surface layers of the silver halide grains of photographic emulsions where the vacant anion site concentration can be affected by a number of external variables (Mitchell 1948).

REFERENCES.

- DIENES, G. J., 1948, *J. Chem. Phys.*, **16**, 620.
 HILSCH, R., and POHL, R. W., 1930, *Z. Physik.*, **64**, 606.
 KOLTHOFF, I. M., and O'BRIEN, A. S., 1939, *J. Chem. Phys.*, **7**, 401.
 LANGER, J., 1943, *J. Chem. Phys.*, **11**, 11.
 MITCHELL, J. W., 1948, *Science et Industries Photographiques*, **19**, 361; 1949, *Phil. Mag.* [7], **40**, 249.
 SEITZ, F., 1946, *Rev. Mod. Phys.*, **18**, 384.
 SIMPSON, J. H., 1949, to be published in *Proc. Roy. Soc. A*.
 STROCK, L., 1934, *Z. Phys. Chem. B*, **25**, 441.
 TUBANDT, C., 1921, *Z. anorg. Chem.*, **115**, 105; 1932, *Wien-Härms Handbuch d. Experimentalphysik*, Leipzig, **12**, 384.
 TUBANDT, C., and LORENZ, E., 1914, *Z. Phys. Chem.*, **87**, 513.
 TUBANDT, C., and EGGERT, S., 1920, *Z. anorg. Chem.*, **110**, 196.

LXIV. *An Improved Schmidt Plate.*

By DOROTHY B. G. HAWKINS, PH.D.,
H. H. Wills Physical Laboratory, University of Bristol*.

[Received March 15, 1949.]

SUMMARY.

In this paper the profile of an improved aspheric plate is obtained which minimizes the off-axis errors of Schmidt cameras over a given field. In the case of monochromatic systems these errors are reduced to two-fifths those of the classical Schmidt if axial stigmatism is retained or to one-third if some spherical aberration is introduced into the system. In the case of chromatic systems a method is given of calculating the equation of the profile which minimizes the overall errors for different spectral ranges over given fields.

THE profile of a Schmidt plate is usually chosen to give axial stigmatism and also to minimize the colour error of the system as in the case of the "classical" Schmidt.

The size of the colour error, however, in Schmidt systems of wide angular fields is small compared with the off-axis errors introduced by the aspheric plate. For instance, it is less than $1/10$ in the case of Schmidt systems with 30° Field. It is shown in this paper that the off-axis errors can be greatly decreased if a different shape of Schmidt plate is used, that is to say, if a different proportion of 2nd to 4th power strength is introduced. This is particularly desirable in systems of high aperture, where the diameter of the circle of confusion due to the off-axis errors may be as large as $1/10$ inch as in the case of an $F/0.6$ system with 10-inch focal length and 30° Field. It may not generally be realized that the wide field in this case only causes a reduction in light-gathering power of some 10 per cent.

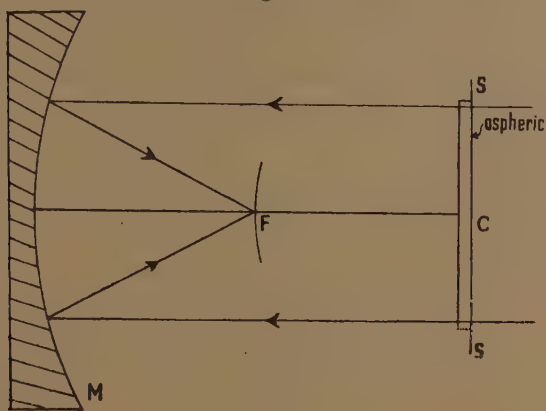
§ 1. PLATE PROFILE WITH MINIMUM OFF-AXIS ERRORS.

In a paper entitled "An Improved Type of Schmidt Camera" (Hawkins and Linfoot, *Monthly Notices, R.A.S.*, Vol. 105, No. 6) a method originally due to Carathéodory was used for obtaining the off-axis aberrations resulting from a Schmidt plate placed at the centre of curvature of

* Communicated by Dr. C. R. Burch, F.R.S.

the mirror in a system working at infinity (see fig. 1). The aberrations at an angular distance ϕ from the centre of the field were calculated by considering the effect on rays traced out from F, of tilting the plate together with the aperture-stop S through an angle ϕ about an axis through C in the plane of the aspheric surface. Cartesian coordinates (x, y)

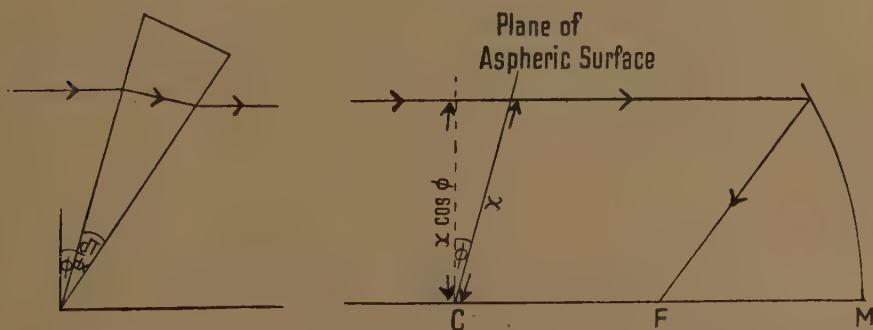
Fig. 1.



Schmidt system.

are taken in the plane of the aperture-stop with origin at its centre C, and the plate is tilted at an angle ϕ about the axis Oy. The deviations caused by the aspheric surface of the untilted plate produce an error-free axial image, and it is the changes which they undergo when the plate is tilted that give rise to the off-axis aberrations. Two main factors operate to produce these changes. The first arises from the fact that the plate

Fig. 2.



(a) "Tilted prism" effect.

(b) "Shift" effect.

is now traversed obliquely by the rays and this increases the angular deviations associated with each point on the plate surface (tilted prism effect: see fig. 2 a). The second from the fact that a point on the front face of the meniscus, which now receives light from the point (x, y) on

the aspheric surface, formerly received it from the point $(x \cos \phi, y)$ (shift effect : see fig. 2 b). When $T(x, y)$ is a measure of the plate thickness the expression

$$\xi + i\eta = (n-1) \sin^2 \phi \left[\frac{n+1}{2n} \frac{\partial T}{\partial x} + \frac{x}{2} \frac{\partial^2 T}{\partial x^2} + \frac{i}{2n} \frac{\partial T}{\partial y} + i \frac{x}{2} \frac{\partial^2 T}{\partial y^2} \right] \quad (1)$$

is found to give to a close approximation the angular aberrations in the field surface of the ray which enters the system at an angle ϕ with the x -axis and which passes through the point (x, y) of the aperture.

The present paper is based on the further consideration that with plate profiles of a certain shape the contribution of the tilted prism effect to the off-axis aberrations is of the opposite sign to that of the shift effect. The size of the off-axis aberrations is thus decreased. It is therefore possible to obtain a particular profile which will make the maximum values of ξ and η as small as possible.

Let the required profile be given by the equation

$$K(n-1)T = -A(x^2 + y^2) + (x^2 + y^2)^2, \quad (2)$$

where the value of A is to be determined. ($K = 4R^3$, where R is the radius of curvature of the mirror. Powers of $x^6 + \dots$ are neglected for the moment.)

Now it can be proved that in axially stigmatic systems where the field surface is taken strictly concentric with the mirror (*i. e.* the usual procedure) the component in the x -direction of the maximum angular aberration of rays passing through any plane $x = ky$ is always greatest when $k = 0$, *i. e.* that the maximum aberration in the x -direction arises from rays passing through the Meridional-plane. Further, that this aberration is always greater than the maximum aberration in the y -direction, therefore we need only minimize the maximum off-axis errors in the Meridional-plane in order to minimize the off-axis errors over the whole of the aperture.

Also since the off-axis aberrations of any ray through the point (x, y) can be expressed to within a sufficient degree of accuracy as the product of the square of the off-axis angle and a plate profile term ξ' , which is independent of ϕ , the plate profile which minimizes the maximum value of ξ' will minimize the off-axis errors all over the field.

From (1) it follows that

$$\xi = \frac{\sin^2 \phi}{K} \xi',$$

where

$$\xi' = \frac{n+1}{2n} (n-1) \frac{\partial T}{\partial x} + \frac{x}{2} (n-1) \frac{\partial^2 T}{\partial x^2},$$

and substituting for values of $\partial T / \partial x$ and $\partial^2 T / \partial x^2$ obtained from equation (2) we have

$$\xi' = -\frac{(2n+1)}{n} Ax + \frac{2(4n+1)}{n} x^3.$$

The value of x which gives a maximum value of ξ' for $0 < x < 1$ is obtained from the equation $\partial \xi' / \partial x = 0$. This maximum value of ξ' will

be minimized when it is made equal and opposite to the value of ξ' at the edge of the plate. A is then found to have the value 2.6. (The effect of different values of n on A is small and can be neglected. For instance, for $n=1.5$, $A=2.62$; for $n=1.6$, $A=2.64$.) Substituting this value for A in equation (2) we obtain $K(n-1)T=-2.6x^2+x^4$ for the equation of the profile of a Schmidt plate giving minimum off-axis error spreads (see fig. 3 a). The profile $K(n-1)T=-1.5x^2+x^4$ of a classical Schmidt plate is shown for comparison (fig. 3 b).

Fig. 3.



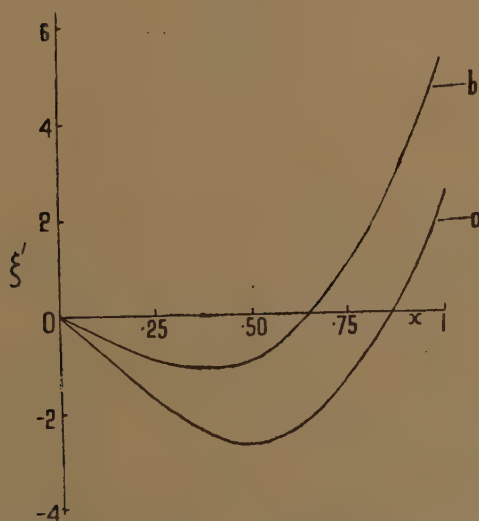
(a) Improved plate-profile (for monochromatic systems).

$$[K(n-1)T=-2.6x^2+x^4.]$$

(b) Classical Schmidt plate-profile.

$$[K(n-1)T=-1.5x^2+x^4.]$$

Fig. 4.



Comparative sizes of off-axis image errors for improved plate (a) and classical Schmidt plate (b).

Fig. 4 shows that not only are the off-axis error spreads reduced by 55 per cent at the edge of the aperture by plate (a) but also that the point at which the off-axis errors are zero now falls on a zone nearer the edge of the aperture, *i. e.* one which makes a larger contribution to the illumination of the image.

Fig. 5 shows the size and shape of the off-axis errors due to the edge zones (*i. e.* the maximum errors) for both the improved plate and the

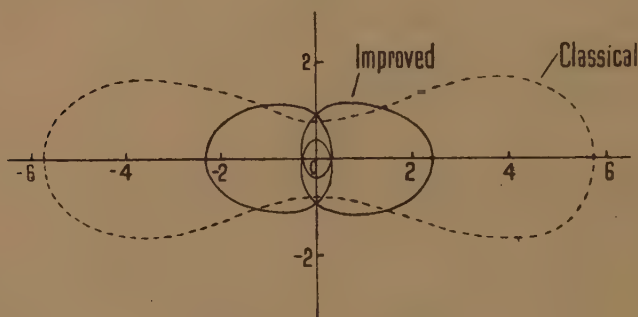
classical Schmidt. The comparisons hold in monochromatic systems for all apertures and for all angular fields. The question of colour error is dealt with later.

The question now arises for what plate-strength does the foregoing analysis hold. In the first place it is of course impossible to deal with systems of the order of $F/1$ and above without reference to terms in x^6 and higher powers. But in the case of axially stigmatic systems of apertures $F/2$ and below the aberration-contribution of terms in x^6 is very small in comparison with that of terms in x^4 and x^2 and does not appreciably affect the best choice of A .

In the case of an $F/1$ system, if the effect of the 6th power term is included and the plate profile is taken as $K(n-1)T = -Ax^2 + x^4 + x^6/2R^2$, the value of A becomes 2.8 instead of 2.6.

There is a second correction which demands consideration in systems of sufficiently short focal ratio; it does not affect the value of A , but

Fig. 5.



Off-axis image spread.

[On-axial image spread nil in both cases.]

Comparison of size of maximum errors for improved plate and classical Schmidt plate for monochromatic systems.

does affect the size of the off-axis image. It is due to the fact that for off-axis rays the plate is no longer tilted in an almost parallel beam. In a classical Schmidt system of $F/1$ working at infinity, rays at the edge of the aperture make an angle of some 14 mins. of arc with the horizontal. This departure from parallelism introduces a higher order comatic error of only 3 per cent of the total off-axis error for an $F/1$ system and of 15 per cent in the case of an $F/0.6$ system. It is found by analysis that this higher order coma can be reduced to about half by moving the aperture-stop a very short distance along the axis towards the mirror ($0.0075R$ in the case of an $F/0.6$).

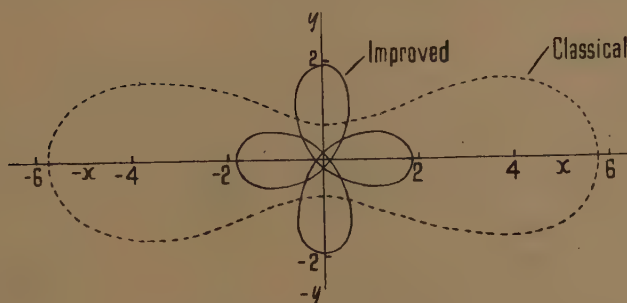
§ 2. INTRODUCTION OF AXIAL SPHERICAL ABERRATION.

A further decrease in the off-axis errors over the whole field can be obtained by introducing a small amount of axial spherical aberration into the system, in other words, by making the Schmidt plate very slightly

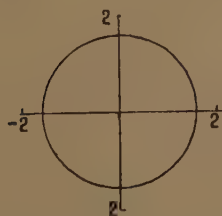
weaker than it would be for axial stigmatism. The amount of spherical aberration needed varies with the field, but is very small even in the case of wide-angled systems. To the degree of accuracy required in practical designing we can take it that the profile which will minimize meridional error spreads will give a close approximation to the minimization of the overall error spreads. Let the equation of this profile be $K(n-1)T = -Ax^2 + x^4(1+a)$. The aberration ξ is now replaced by $\xi + 4ax^3 = (\sin^2 \phi/K)[\xi' + 4a'x^3]$, where $a' = Ka/\sin^2 \phi$ and ϕ denotes the semi-angle of the field of view. In order to find values of A and a' which will minimize the above expressions, we observe that for any equation of the form $S = Bx + x^3$ if B is so chosen that $S_{x=1} = -S_{\max}$ the maximum value of S is at $x = \frac{1}{2}$. Hence the required values of A and a' are obtained from the solution to the equations,

$$[\xi' + 4a'x^3]_{x=1} = -[\xi' + 4a'x^3]_{x=\frac{1}{2}} = [-4a'x^3]_{x=1}. \quad (3)$$

Fig. 6.



(a) Off-axis image spread.



(b) Axial image spread. (Nil in case of classical.)

Comparison of size of maximum errors for monochromatic systems.

It is found that the required values are $A=2.1$, $a'=0.46$. This gives $K(n-1)T = -2.1x^2 + x^4(1-0.46 \sin^2 \phi)$ for the equation of the plate profile. (The effect of different values of ϕ on the values of A and a is exceedingly small and can be neglected.)

In fig. 6 the size of the off-axis image of a system with the improved plate (plus spherical aberration) is shown against that of a system with a

classical Schmidt plate. It will be seen that the errors of the former have been reduced to one-third those of the latter. This comparison holds to a close approximation for all fields and for all apertures in monochromatic systems.

§ 3. CHROMATISM.

The above considerations still apply in the case of chromatic systems. Here again, by correct choice of the value of A, a considerable reduction in the size of the off-axis error spreads can be obtained. Ignoring for the moment the question of the introduction of spherical aberration, the required value of A is obtained from the equation

$$\left[\xi' + \frac{(n_c - n_d)}{\sin^2 \phi} \frac{\partial T}{\partial x} \right]_{x=1} = - \left[\xi' + \frac{(n_F - n_d)}{\sin^2 \phi} \frac{\partial T}{\partial x} \right]_{x=\frac{1}{2}} \quad \dots \quad (4)$$

in the case of a system giving axial stigmatism in *d*-light and working over the range C-light to F-light. This equation is for values of A ≥ 2.0.

TABLE I.

Colour range	ϕ	A	Size of off-axis error spread. K ξ (in radians)	Percentage improvement over classical Schmidt	Axial colour $x=1$	spread $x=0.6$
C-F	15°	2.5	0.18	53%	-0.014	-0.031
	9°	2.4	0.076	47%	-0.011	-0.028
	6°	2.1	0.041	43%	-0.004	-0.024
C-G	15°	2.4	0.19	49%	-0.025	-0.061
	9°	2.1	0.093	42%	-0.005	-0.049
	6°	1.8	0.055	38%	+0.008	-0.041

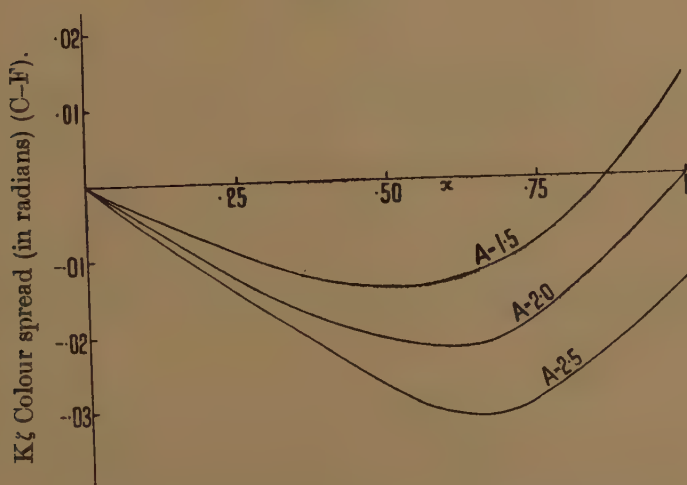
For apertures of F/1 and above the corrections discussed in section 1 must be made. The figures for the percentage improvement over the classical Schmidt still hold.

“The value A=2.0 represents a profile with neutral zone at the edge of the plate.” For A<2.0 the factor $(n_c - n_d)$ must be replaced by the factor $(n_F - n_d)$, since the error in F-light is now greater than that in C-light at $x=1$. If a larger spectral range is required, such as C-light to G-light, then the factor $(n_F - n_d)$ must of course be replaced by $(n_G - n_d)$. The value of A varies with ϕ and Table I. shows the corresponding values of A for different values of ϕ , together with the size of the total off-axis errors (ξ) and the percentage improvement in each case over the classical Schmidt. Mean values of $(n_F - n_d)$, $(n_c - n_d)$ and $(n_G - n_d)$ are taken, viz. : $(n_F - n_d)=0.007$, $(n_c - n_d)=-0.003$ and $(n_G - n_d)=0.015$. The effect of small variations in these values can be neglected.

It will be appreciated that in general the longer the focal ratio the smaller the value of ϕ (since the central obstruction will cause a greater reduction in light-gathering power). The importance of the colour error as compared with the off-axis error is therefore increased for long focal ratios. For a system with 4° Field A becomes 1.6.

The difference in the axial colour spread over the aperture for various values of A is shown in fig. 7. Where ϕ is small, say 6° the alteration in the value of A causes the maximum colour error, which falls on the zone $x=0.6$ approximately, to be $1\frac{1}{2}$ times as great as in the classical Schmidt. For the edge zones, however, which contribute most of the light it is less than in the classical Schmidt. When ϕ is increased to 15° and A becomes 2.5 the maximum colour error increases to more than

Fig. 7.



Comparative axial colour spreads for various values of A.

twice that of the classical Schmidt but for such values of ϕ the colour error is only a very small proportion of the total errors of the system (see fig. 8). This comparison holds for all spectral ranges.

A further improvement in the size of the error spreads can be obtained as before by the introduction of a small amount of spherical aberration. The equation of the plate profile becomes $K(n_d-1)T = -Ax^2 + x^4(1+a)$ and values of A and a' in this case are obtained from equations

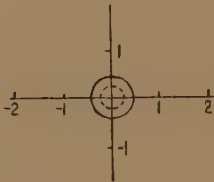
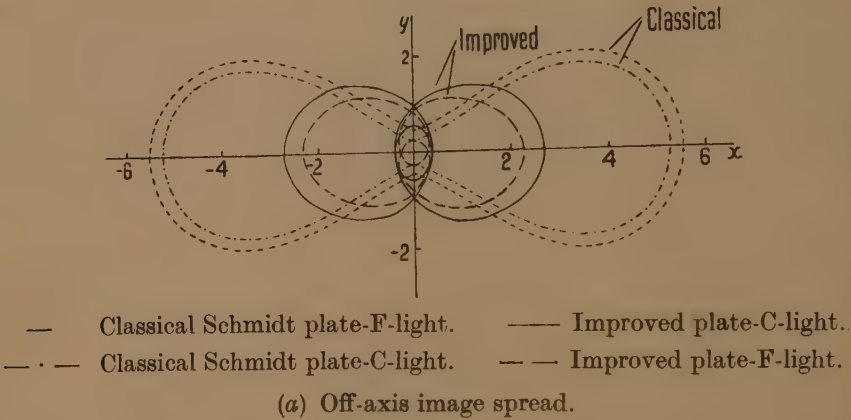
$$\left[\xi' + 4a'x^3 + \frac{(n_c - n_d)}{\sin^2 \phi} \frac{\partial T}{\partial x} \right]_{x=1} = - \left[\xi' + \frac{(n_F - n_d)}{\sin^2 \phi} \frac{\partial T}{\partial x} + 4a'x^3 \right]_{x=\frac{1}{2}} \quad (5)$$

$$(4a'x^3)_{x=1} + \frac{(n_F - n_d)}{\sin^2 \phi} \left(\frac{\partial T}{\partial x} \right)_{x=0.6} = - \left[\xi' + 4a'x^3 + \frac{(n_c - n_d)}{\sin^2 \phi} \frac{\partial T}{\partial x} \right]_{x=1} \quad (6)$$

If A is < 2.0 the factor $(n_c - n_d)$ must be replaced by the factor $(n_F - n_d)$. In equation (6) the maximum axial error is taken as the maximum spherical aberration together with the maximum colour error. This ensures that

the maximum off-axis error is slightly larger than the maximum axial error. Fig. 9 shows the comparative sizes of the maximum off-axis image spreads of a system with 30° Field. The points in the diagram are in F- or C-light according to which gives the larger image.

Fig. 8.



(b) Axial image spread.

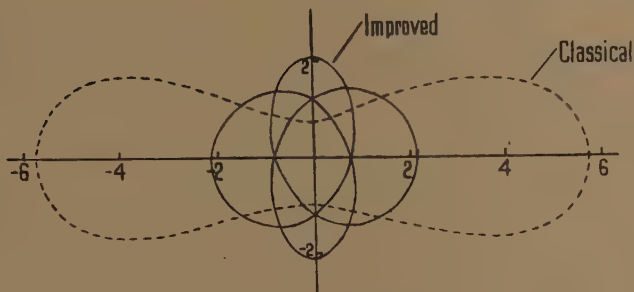
Comparison of maximum errors for improved plate and classical Schmidt plate for chromatic system with 30° Field.

TABLE II.

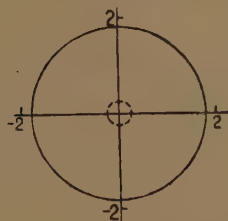
Colour range	ϕ	A	a'	Size of off-axis error spread. $K\zeta$ (in radians)	Percentage improvement over classical Schmidt	Axial error. $K\theta$ (in radians)	
						$x=1$	$x=0.6$
C-F	15°	2.04	-0.45	0.14	67%	0.12	0.014
	6°	1.87	-0.37	0.035	51%	0.023	0.014
C-G	15°	2.03	-0.40	0.16	60%	0.11	0.12
	6°	1.74	-0.34	0.052	41%	0.030	0.022

Values of A and a' , together with the percentage improvement in the size of the off-axis image spreads, are shown in Table II. for the spectral ranges C- to F-light and C- to G-light. The axial error θ is the sum of the introduced spherical aberration and the axial colour error.

Fig. 9.



(a) Off-axis image spread.



(b) Axial image spread.

Comparison of maximum errors for chromatic system (C-F) with 30° Field when spherical aberration introduced into improved system.

ACKNOWLEDGMENT.

I wish to express my thanks to Dr. C. R. Burch, F.R.S., Dr. E. H. Linfoot and Dr. E. Wolf for helpful conversations during the course of this work.

LXV. Notices of New Books and Periodicals received.

Les Fondements Mathématiques de la Mécanique Quantique. Par J. VON NEUMANN. Traduit par A. Proca. (Paris: Librairie Felix Alcan, 108 Boulevard, Saint-Germaine, 1946.)

THIS work is a translation, by a well known theoretical physicist, of von Neumann's great book: *Mathematische Grundlagen der Quantenmechanik*, which was first published about 1932. In recent times (1943) it has been reprinted and published by authority of the United States Alien Property Custodian.

The purpose of the work, as it is described in Proca's translation, is to give *un exposé cohérent homogène et, autant que possible, mathématiquement rigoureux*. Great importance is attached to questions of principle and of interpretation and this indeed is one of the great merits of the book. The application of quantum mechanical methods to particular problems is not dealt with.

The book suffers from being compressed into too small a space. Its intelligibility and its value might have been enormously enhanced had the author used at least twice as many pages. The translation impresses me as very good and will probably help many who might find the German original a little obscure. The leaves of the volume have to be cut, as is usually the case with French mathematical books. There is no index; but a very complete *Table des Matières*.

Von Neumann prefers to work with the operators themselves rather than with the corresponding matrices. Paradoxically his book appears to me to suffer from his constant concern with the *exigences de la rigueur mathématique* and in the early part of it he takes Dirac severely to task for failing to reach the *minimum habituel requis en Physique Théorique*. In particular he criticizes (as others have done) the δ function and here I feel that the simple fact is that he has not rightly appreciated the nature of the δ function. It is more like an operative symbol than a function and only occurs in expressions which represent the limits of sequences of integrals. Von Neumann himself compares it (*i.e.* $\delta(q)$) with

$$\lim_{a \rightarrow \infty} \sqrt{(a/\pi)} \exp(-a^2 q^2),$$

but in treating the integrals involving it he ignores the limit aspect and thus of course makes nonsense of it. These integrals, it may be added, and the δ function involved in them, do not differ very materially from corresponding things in a great paper by Lejeune Dirichlet (1837).

On page 44 there is a characteristic paragraph. The author is proving that when $(f, f) = 0$ it necessarily follows the $f \equiv 0$. It should be explained that (f, f) means the integral $\int \dots \int f^2 dq_1 \dots dq_k$ and f^2 , if not zero, is real and positive over the domain of integration. The unsophisticated physicist would establish quite simply and (I think) soundly that $f \equiv 0$ everywhere in the domain. Not so von Neumann. He has given his allegiance to Lebesgue's form of integration and therefore must permit f to differ from zero in an *ensemble de points* $q_1 \dots q_k$ *de mesure nulle*. So he adopts the *convention* that f may be regarded as equivalent to zero ($f \equiv 0$). It should not be overlooked that an *ensemble de mesure nulle* includes all rational values of $q_1 \dots q_k$ whatsoever! A chapter is devoted to Hilbert's space which, very briefly and roughly, is the space or continuum, suggested by the analogy between orthogonal vectors on the one hand and orthogonal functions on the other. There are chapters on quantum statistics, including the *Théorie de la lumière*; the axiomatic development of the theory and on the measuring process.

The most interesting part of the book is the discussion of the question as to whether there is causality in the physical world. Von Neumann is quite convinced there is not—except such as is due to large numbers, but perhaps the term “causality” should not be used for what is only a macroscopic simulation. He contends that the quantum theory is in logical contradiction with the causality principle; that it is incompatible with the existence of hidden parameters. This may conceivably be true; but I am not convinced. As von Neumann himself points out there is strongly marked causality even among quantum laws, as witness

$$\frac{\partial \phi}{\partial t} = -\frac{2\pi i}{h} H\phi.$$

Acausal changes, if I have understood the author rightly, are connected (only?) with the intervention of the observer and this aspect, I feel has been discussed inadequately.

It is a most valuable book, which I have read and am still studying with much profit; but it should be read critically, as no doubt the author himself would agree.

W. WILSON.

Theory of Groups and its Application to Physical Problems. By S. BHAGAVANTAM and T. VENKATARAYUDU. [Pp. xi+234.] (Waltair, India: Andhara University.) 20 rupees.

GROUP theory first found its way into physics in connection with the discussion of crystal symmetry: more recently it has been extensively applied to quantum-mechanical problems. In the present volume these two fields are interwoven.

The necessary amount of the theory of groups and their matrix representation is developed *ab initio*. The orthogonality relations of group characters are derived by the use of the Frobenius algebra of the group: a more straightforward if less elegant method might have been preferable. The authors give the impression that the group of 2×2 unimodular unitary matrices forms a representation of both the three-dimensional rotation group and the Lorentz group. For the latter the limitation to unitary matrices must be dropped.

The choice of applications is strongly influenced by the authors' special interests. The well-known applications to atomic and molecular spectra are sketched, but the use of group theory in dealing with molecular resonance-energies is not mentioned. In compensation the treatment of Raman and infra-red spectra is very full, covering both molecules and crystals. A brief account of lattice space-groups is used to discuss the vibrations of a crystal from the point of view of Raman's theory. Detailed treatment of particular molecules and crystals illustrates the general argument. The book concludes with character tables for the point groups.

J. D. E.

Supersonic Flow and Shock Waves. By R. COURANT and K. O. FRIEDRICHS. [Pp. 464.] (Interscience Publishers Ltd.) Price 42s.

IN recent years, particularly during the late war, gas dynamics has attracted a large number of research workers, and the vast number of published research papers, and reports issued by various official agencies, is out of all proportion to the available text-book literature, particularly text-books in English. In such circumstances, it needs courageous authorship to produce, in text-book form, a fundamental framework into which so many diverse and incomplete researches can be fitted to form something like a composite whole. This volume, by Professors Courant and Friedrichs, is a very worthy attempt to provide for such a long-felt need, and marks a vast improvement on their original draft, published during the war as a report of the U.S.A. Office of Scientific Research and Development, which came to be known as 'The Shock Wave Manual'.

The book attempts to present a systematic theory of non-linear wave propagation, and the emphasis is almost entirely on gas dynamics. After a full discussion of the thermodynamical concepts involved, and the mathematics of hyperbolic equations in two independent variables, the two main chapters, which occupy most of the book, are devoted to unsteady flow on one dimension, and isentropic irrotational steady plane flow. In both cases the essential rôle of the characteristics in continuous flow and the character of shock discontinuities are fully developed from first principles. Considerable space is allotted to the discussion of shock interactions, including irregular or Mach reflection of shocks. The later chapters are devoted to flow in nozzles and jets, steady flow with cylindrical symmetry, conical flow and spherical waves. In addition, space is found to include such topics as wave propagation in elastic-plastic materials, detonation and deflagration waves, and the shallow-water analogue of non-linear wave motion.

The authors have not attempted to give a comprehensive survey of all the literature, but they give a list of other works which contain extensive bibliographies, a list of surveying reports, and then a very useful and up-to-date selection of pertinent references, arranged in order, chapter by chapter, as they are referred to in the relevant paragraphs in the text.

Throughout the book the authors emphasize the incompleteness of present knowledge and the lack of perfection, as a mathematical theory, which the non-linear character of the basic equations entails. For this reason they strive continually to stress the gaps and uncertainties and to maintain as general a viewpoint as possible.

With such a wide range to cover, however, it is inevitable that there should be gaps here and there, where practical considerations have limited the advances made in special types of problems. Thus, although much attention is paid to the problem of a shock wave running into a corner, and its subsequent reflection, the complementary problem of a shock wave being diffracted round a corner is not mentioned. Again, for instance, the method of characteristics in more than two independent variables is dismissed as not yielding an essential simplification. On the contrary, many of the most important practical problems in gas dynamics are concerned with unsteady motion in two space-variables, in which case a reduction to characteristic form yields equations involving derivatives in only two directions. Some discussion of the geometry of characteristic cones and surfaces in this type of problem, and of possible methods of numerical computation analogous to those for flow with two independent variables, might well have been included.

The book is excellently composed from the point of view of arrangement and continuity, though the authors have clearly had formidable difficulties over notation, and, in places, a better and perhaps simpler choice of wording would improve the lucidity of the argument or explanation.

The text is remarkably free from misprints, but some of the figures need a little more scrutiny. There is an amusing draughtsman's error in figures 14 and 15, pp. 274, 276—amusing because it is such a complete negation of everything in the text. In both these figures, a uniform supersonic flow starts to turn round a corner before encountering the simple wave set up by the corner!

But no such minor shortcomings can detract from the real merits of this excellent work, which definitely surpasses previous text-books on the subject in comprehensiveness and thoroughness of treatment of the fundamental concepts. It will doubtless be, as the authors express hope, most helpful to engineers, physicists, and mathematicians alike, and it is safe to say that no one who has interests in the field of gas dynamics, or indeed in any other branch of non-linear wave propagation, can afford to be without access to it. C.K.T.

Progress of Theoretical Physics, Vol. 3, Nos. 1, 2, 1948. Edited by H. YUKAWA. (Akitatya Co. Ltd., Osaka, Kyoto, Japan.) Price 120 yen per issue.

THEORETICAL physicists everywhere will welcome wholeheartedly the inception of a specialist journal devoted almost entirely to theoretical physics of a fundamental character, especially since it contains the work of the active and fruitful school of Japanese theoreticians lead by Yukawa. In the two quarterly issues under review here, the majority of papers are concerned with field theories of the elementary particles. These include the later papers of the series entitled "On a Relativistically Invariant Formulation of the Quantum Theory of Wave Fields" by Tomonaga and his co-workers, who have developed, independently of the later American authors, the subtraction processes which enable avoidance of the many divergences of field-theoretical calculations, so long a spectre in the path of progress to further knowledge of fundamental processes.

However, in accord with the expressed policy of the journal, these issues contain also papers on more well-established fields of quantum theory. For example there are papers on Ferromagnetism, on Interval Formulæ for Atomic Multiplets and on the Factorization Method for solution of eigenvalue problems, as well as discussion of experimental results with an immediate bearing on contemporary theoretical physics.

Unfortunately, the text contains much incorrect spelling, many misprints and some rather un-English constructions, often of elusive interpretation, which may cause confusion to many readers; yet 'Progress in Theoretical Physics' will, undoubtedly and with good reason, be much sought after by all those interested in the current progress of fundamental physical theory.

It provides an interesting example of the way in which a journal may specialize on one branch of physics; we have not in this country or in the U.S.A. any similar journal specializing on theory alone.

Two Lectures. By W. HEISENBERG.

THESE lectures were delivered at the Cavendish Laboratory during Professor Heisenberg's recent visit to Cambridge. They give an account of the author's views on two of the most fundamental problems in theoretical physics to-day.

The first lecture discusses the theory of the elementary particles, a part of physics in which the basic laws are still unknown. Professor Heisenberg develops his view that a complete description of the elementary particles will only be achieved by means of a formalism which comprises all the particles together, in a way similar to that in which a complex system in unrelativistic quantum theory is described by a Hamiltonian which refers to all the constituent nuclei and electrons. A discussion is given of the probable rôles played in the future theory by the author's S-matrix theory and by a new universal constant of the dimensions of a length.

The second lecture deals with the theory of superconductivity and analyses the physical aspects of the author's recent attempt to interpret the phenomenon in terms of the quantum theory of electrons in metals. The new theory is based on the idea that at low temperatures a condensation process can take place among the free electrons in a metal. The consequences of this assumption are examined in detail, and an attempt is made to use them to obtain a consistent description of the observed properties of superconductors.

The ideas put forward in these lectures represent a most valuable contribution towards our insight into these extremely difficult problems, even if the final solutions should be found along lines somewhat different from those laid down here.

E. H. S.

Natural Philosophy of Cause and Chance. By MAX BORN. (London: Geoffrey Cumberlege.) Price 17s. 6d. net.

THIS book is divided into two sections, a general argument in ten chapters extending over 128 pages, and an appendix of amplification and justification of 81 pages, followed by a good index. References to original publications are given more freely than is usual in a semi-popular book of this type.

The argument begins with a contrast of causality with determinism. Accepting antecedence and contiguity as principal elements of the causal relation, Born points out the symmetry with regard to time reversal of the equations of classical mechanics and electrodynamics, a unique direction of time being found only in the laws of variation of the probability distributions describing incompletely specified systems. A discussion of statistical mechanics leads to an examination of the use of probabilities in quantum theory, based on the modern concept of the state of a physical system as the epitome of its history, giving the probability distribution of the results of all measurements which may be made upon it. Born, however, emphasizes that the essential concept of causality is retained in the evolution of the functions determining these probabilities, governed by the differential equations of quantum mechanics. All that is lost is the metaphysical (in the worst sense) idea of mechanical determinism.

The book is not a Natural Philosophy, but an admirable clearing of the undergrowth, preparatory to laying the foundations for one.

G. W.

Calcolo Tensoriale e Applicazioni. By B. FINZI and M. PASTORI. [Pp. vii+427.] (Bologna: Nicola Zanichelli, 1949.) Lire 2,000.

THE tensor calculus and its applications owe much to the Italian mathematicians. The English work which most nearly covers the same ground as the present volume is McConnell's *Applications of the Absolute Differential Calculus*, and McConnell himself studied in Rome. Finzi and Pastori work to a more advanced level than McConnell, treating non-Riemannian spaces and general relativity. The first hundred pages are devoted to vector and tensor algebra, and the next hundred to vector homographies, and tensor fields in Euclidean and non-Euclidean spaces. There follow chapters on the geometry of surfaces and of Riemannian spaces. Physical applications occupy the last 140 pages, with chapters on the mechanics of continua, the electromagnetic field, and relativity. A good index and tables of formulæ at the ends of chapters make it a useful reference work.

F. R. N.

Science and its Background. By H. D. ANTHONY. [Pp. ix+304.] (Macmillan.) Price 10s. 6d.

Dr. ANTHONY has attempted something much more difficult than just a short history of science and in many ways more valuable. He introduces the development of science as part of the general process of development of civilization. To simplify and concentrate his narrative only the greater discoverers are introduced. The method is a bit hard on the lesser lights, as individuals, but in Dr. Anthony's hands produces no serious distortions. It is better in its results than the attempt to put in everybody, even if he is only a name and a date.

The book, which is illustrated with skill and originality, should fulfil the needs of the Sixth Forms of schools excellently, and also those of the ordinary reader with no special scientific knowledge. The author seems very reliable; in the course of a careful reading the reviewer has found two small verbal slips (pp. 223 and 224) and two passages to quarrel with. The relation between James Watt and Joseph Black is described (p. 192) as though a casual acquaintance; but Black was a lecturer at Glasgow University when Watt established his workshop there, and the two were closely associated. Watt's steam engine depends on Black's discovery of the latent heat of steam and is the first invention that is real applied science, not just common sense and good guesswork. The other point concerns Dalton and the Atomic Theory, admittedly a tangled affair. Dr. Anthony hardly makes it clear enough that Dalton's discovery came from his study of gases and the Kinetic Theory already stated in outline by Newton. Moreover, Dalton's own thinking was confused in some respects and these confusions were not finally cleared up until Avogadro's ideas were taken up after a lapse of some years. The whole story is a very revealing example of the way true insight and error may go together.

A. D. R.

An Elementary Handbook of Logic. By JOHN J. TOOHEY, S.J. 3rd Edition. [Pp. xiv+194.] (New York and London, Appleton.) No price.

THIS is a thorough, systematic, but not too lengthy, exposition of Aristotelian formal logic, with copious examples and exercises to help the student. The treatment is orthodox. On some few points the author differs from Dr. Keynes, and he treats the Fourth Figure of the syllogism with the contempt it deserves. The difficult marginal subjects of Predicables and Categories, of Classification and Definition are dealt with briefly, perhaps too briefly. The chapter on Fallacies is more adequate than in most textbooks.

It is a pity that nothing is said in Introduction or Appendix about the relation between this older formal logic and the new symbolic logic derived from Boole. The most elementary student should know that both exist and that there are sound reasons for giving him the old before the new.

A. D. R.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

LXVI. *Capture of Negative Mesons.*

By R. HUBY,

Department of Theoretical Physics, University of Liverpool*.

[Received May 2, 1949.]

ABSTRACT.

The time taken for a negative meson, stopping in a solid, to reach its lowest Bohr orbit about a nucleus, is discussed, with special reference to the possibility of "trapping" for a long time in a bound orbit in insulators. It is shown that there are more ways in which this can happen than are dealt with in Fermi and Teller's discussion; and an approximate calculation is made of the time spent in a trap. The results are consistent with the experimental observations on π and μ mesons. For heavy mesons (mass $\sim 900 m_e$), the time of approach to a nucleus is found to be normally $10^{-13} - 10^{-12}$ sec., but trapping of mesons may be expected in certain circumstances for times up to 10^{-5} sec.

§1. INTRODUCTION.

FOR the interpretation of experiments on the fate of negative mesons brought to rest in matter, it is important to have an estimate of the time which the meson, starting say with a small positive energy, requires to fall into its lowest Bohr orbit about a nucleus, provided that spontaneous decay does not happen first. This time will be termed the "time of approach" τ . The theoretical calculation of it was discussed in detail for a meson of mass $200 m_e$ by Fermi, Teller and Weisskopf (1947), and Fermi and Teller (1947) (referred to hereafter as F.T.)†, who predicted that in condensed matter τ should be of the order of 10^{-13} sec., except possibly in insulators for which $Z < 6$. This is consistent with the following experimental information yielding an upper limit for τ :—

(a) From the observed high probability in photographic plates that a slow meson ejected from a nuclear disintegration will itself produce a disintegration, Occhialini and Powell (1948) conclude that the time of approach for these mesons (believed to be of mass approximately $300 m_e$) is much less than the decay time, approximately 10^{-8} sec. This will apply to capturing atoms of $Z \geq 6$.

* Communicated by the Author.

† A part of these calculations has been confirmed by Mott (1949). *Note added in proof.*—It is learned that the problem has also been discussed by Ferretti (*Nuovo. lim.* 5, 325 (1948)).

(b) Many experiments in which sea-level mesons (mass $200 m_e$) disappear without decaying in substances of high atomic number, show that the time of approach in these media is much less than $2 \cdot 10^{-6}$ sec. Further, the measurements of Ticho (1948) with absorbers of Z between 10 and 15 give capture mean times for negative mesons of the order of 10^{-6} sec., and Ticho shows that the dependence of these times on Z agrees with Wheeler's (1947) theoretical prediction, on the assumption that the interaction of the mesons with nucleons is rate-determining, *i.e.* the time of approach is much shorter than 10^{-6} sec.

However, the theory of the time of approach is felt to be worth further consideration, for the following reasons :—

(a) Possible loopholes in the arguments of F.T. In insulators there is a possibility of the existence of "traps" for the meson, *i.e.* bound orbits in which its loss-rate is much less than that given by F.T. These authors deal with the question of traps, but conclude that they do not exist for media of $Z \geq 6$. However, that traps can in fact exist in these media is argued in §§ 2 and 3 below.

An alternative way of dealing with energy loss for mesons of small positive energy in insulators is discussed by Rosenberg in the preceding paper, and this part of the meson's motion will therefore not be dealt with further here. The general problem has also been considered by Fröhlich *et al.* (1948).

(b) Actual loss-rate in traps. If traps are important, it is desirable to have an estimate of the actual loss-rate in them. An attempt to calculate the loss-rate for a very slowly-moving meson in a bound orbit is made in the present paper.

(c) Recent indications of the existence of heavy mesons (mass $\sim 900 m_e$) makes it desirable to have an estimate of the time of approach for these also if they should exist. This is dealt with in § 4.

The loss-rate for a very slowly moving meson in a bound orbit is found to depend so exceedingly sensitively on various parameters, that no reliable figures can be given. However, the sense of the variation with such parameters as Z of the capturing atom, and the mass of the meson, can be ascertained; and an idea can be obtained of the circumstances in which trapping might be important for light and heavy mesons respectively.

§ 2. TRAPPING OF MESONS IN GENERAL.

In a metal F.T. deal with the loss of energy from mesons with energy over a wide range, by considering collisions with individual electrons in the Fermi-gas. They show that, from the meson's free state down to bound orbits well inside the electronic K shell, the loss-rate is approximately given by :

$$-dW/dt \sim T/t_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where T is the kinetic energy of the meson, and $t_0 = \mu \hbar^3 / m_e^2 e^4$. (μ —mass of meson.) In individual collisions, the order of magnitude of energy

loss is $m_e v_0 V$, where v_0 is the maximum velocity of the electrons in the Fermi-gas, and V the velocity of the meson.

Where necessary, the value of T used for (1) is a suitable statistical average \bar{T} taken over the volume of the atom [F.T. equation (18)].

In insulators, F.T. use the same method, but loss according to (1) is assumed to occur only provided the loss per collision is greater than the "Brillouin Gap" G . Thus for any energy W , an average \bar{T} can be calculated, excluding contributions from those regions of the atom in which

$$m_e v_0 V < G, \quad (2)$$

and substitution of this \bar{T} in (1) yields the average loss-rate. It is shown that the order of magnitude of the time of approach so obtained is not different from that for metals. However, the use of the average \bar{T} can be quite wrong for individual mesons, and in fact a meson will be "trapped", unable to lose energy, if it falls into a periodic orbit satisfying (2).

This possibility of traps raises three questions :—

- (a) In what regions of the atom can traps exist?
- (b) What is the probability that a meson falls into one?
- (c) What is the actual loss-rate in a trap?

F.T. discuss the first question for a neutral atom, using the Thomas-Fermi approximation, and considering circular orbits of the meson. They show that stable orbits can exist only within a certain radius of the nucleus; while for any given energy (2) is only satisfied outside a certain radius. They find that the likelihood of the overlap required for a trap decreases with increasing Z , and increases with G . However, F.T. appear to have overlooked the possibility that the meson when bound will most probably be moving, not in a neutral atom, but in a positive ion, since in the capture process an electron will probably have been ejected. Now a positive ion possesses more stable orbits than a neutral atom. For a given degree of ionization, a light ion can probably have stable orbits at any radius, while heavy ions will have an outer shell and an inner core where stable orbits can exist. The inner orbits may not be "loss-free" [*i. e.* satisfying (2)], but the outer orbits almost certainly will be. Thus traps can probably exist in insulators of any atomic weight.

Question (b) cannot be answered with any certainty, but that there is a finite probability can be seen as follows. Capture will occur as the result of energy loss in a collision with an electron. If the excitation energy for this electron is, say, G_0 , then the capture process can take place in any region in which

$$m_e v_0 V > G_0, \quad (3)$$

for a meson of practically zero energy. After capture, the meson will be in a trap, if moving on a closed orbit for which $m_e v_0 V < G_1$, where G_1 is the lowest energy required to excite a further electron. Since G_1 is clearly greater than G_0 , the possibility of falling into a trap does exist,

provided only that G_0 and Z are not so large that (i.) condition (3) excludes all the outer shell of stable orbits, and (ii.) there are no traps in the inner core of stable orbits, with respect to transfer of energy G_1 to an electron. This exceptional case seems unlikely for any element, especially as G_0 is probably very small if the capturing atom is initially neutral, owing to the static repulsion exerted by the meson on an electron: indeed G_0 might even vanish, in which case capture could take place anywhere.

Although no estimate of the probability of capture into a trap has been made, it can be said with confidence that it will be larger the lighter the atom.

To answer question (c) an estimate has been made of the rate of transfer of energy to bound electrons by a meson, considered as a classical particle moving in a circular orbit. It is assumed that the meson-electron interaction disturbs the motion of an electron so much more than that of the meson, on account of the mass disparity, that the meson can be considered to persist in rotation with a uniform angular velocity ω ; and the energy transfer can be calculated as the energy change of the electron. A modified first-order perturbation method, which takes account of the nearly adiabatic nature of the electron's motion is used; details are given in Appendix I. In the calculations, as far as possible, an upper limit of the meson loss-rate has been sought throughout. It is found that electrons can be ejected only into states of rather high angular momentum; in fact, with magnetic quantum number greater than approximately M_0 , given by

$$M_0 = \frac{G}{\hbar\omega}, \quad \dots \dots \dots (4)$$

The matrix elements for transitions to such states, and hence the rate of energy transfer, fall off exponentially as M_0 increases. The formula for meson energy loss-rate to the valence electrons in one atom is

$$-\frac{dW}{dt} = \frac{W_H}{\tau_H} \frac{A}{M_0^2 s} e^{-L}, \quad \dots \dots \dots (5)$$

where
$$L = (M_0 - l_0 + \frac{1}{2}) \left(\frac{s^2 - 1}{s} \right) - \pi Z' \left(\frac{W_H s}{G} \right)^{\frac{1}{2}}. \quad \dots \dots \dots (6)$$

W_H is the ground-state energy of the hydrogen atom, and τ_H is the corresponding period of rotation on the Bohr model, its value being 1.5×10^{-16} secs.

A is a number, of order of magnitude unity, depending on the parameters of the initial state of the ejected electron (see equations (34) and (35)). Z' is the effective charge of the field in which the ejected electron moves; it is taken as unity if one electron had already initially been ejected during the meson capture.

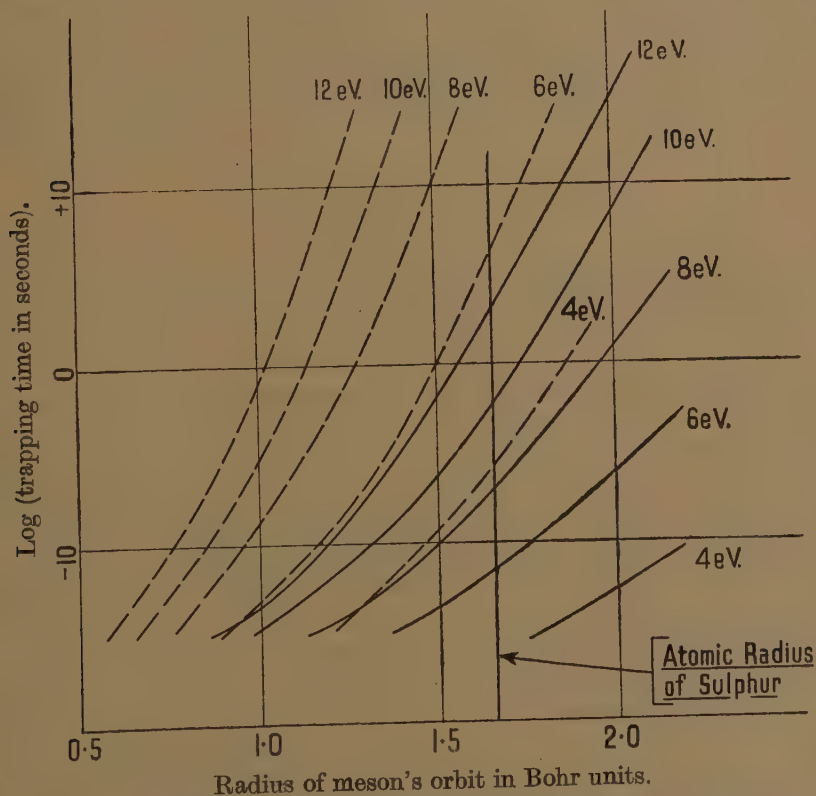
l_0 is the azimuthal quantum number of the initial state. s is the solution of the equation:—

$$s - \ln \frac{(1+s)^2}{s} = 2 \ln M_0 + \ln \left(1 - \frac{\rho_0}{M_0} + \frac{1}{2M_0} \right) - \ln \left(\frac{e}{4} \frac{G}{W_H} \frac{a^2}{a_H^2} \right) \quad \dots \quad (7)$$

where a is the radius of the meson's orbit, and a_H is the Bohr radius. s is very nearly equal to G/\bar{W} , where \bar{W} is the mean energy of the ejected electron.

The rate of loss is governed chiefly by the exponential factor e^{-L} , and is extremely sensitive to small changes in the parameters of the meson and electron orbits. A formula giving the order of magnitude of the time τ_{trap} spent by the meson in any trap has been obtained by estimating this time for a typical condition, and then supposing that the variation is determined solely by L . The result is

$$\tau_{\text{trap}} = 10^{L \log e - 15.6} \text{ sec.} \quad (8)$$



Time spent by a meson in a trap, if Auger loss alone is operative.

Full lines: Meson of mass $200 m_e$

Broken lines: Meson of mass $900 m_e$

The curves are labelled according to the "Brillouin Gap," G , of the insulator.

The typical case taken was that of a $200 m_e$ meson moving in sulphur on the periphery of the atom. The time according to eqn. (8) is plotted in the figure against the radius of the meson's orbit, for various values of G , and two values of the meson mass, viz., 200 and $900 m_e$; these are the only variables on which the time depends in this approximation, if Z' is

assumed always to be unity.* Details of the approximate treatment of the variables in these calculations are given in Appendix II.

The crudity of the approximations is such that the results can only be used either to exclude extreme improbabilities, or to find the sense of variation of the time with the variables concerned. The time increases rapidly with meson mass, and with radius of the meson's orbit, the reason in each case being the diminution of the angular velocity of the meson. The rapid variation with a/a_H means that traps in different parts of the same atom will have very different times. The time also increases with G , which represents the difficulty of removing an electron; and since this is the only important factor depending on the nature of the medium, it follows that the time in various substances depends more on their chemical characteristics than on the atomic weight,

When the time thus obtained is very long, radiation loss of energy becomes rate-determining. The order of magnitude of the time spent in a trap, if radiation alone were operative, can be obtained by classical theory

$$\tau_{\text{rad}} \sim \frac{\mu^2 c^3 a^3}{4e^4 Z(a)}, \quad (9)$$

where $Z(a)$ is the charge within the radius a . This also increases with meson mass and radius, but much more slowly than the loss to electrons.

3. MESONS OF MASS 200–300 m_e .

In the following, a meson mass of $200m_e$ will be assumed, but the result for a $300m_e$ mass would not be very different. F.T. find that in metals the time of approach τ is about 10^{-13} sec. and of the same order of magnitude in insulators, if there are no traps. They find that there are no traps for insulators with $Z \geq 6$, on the assumption that $G \leq 7eV.$, as in these circumstances the Thomas Fermi model for neutral atoms yields no stable orbits satisfying (2) (or more strictly the condition $m_e v_0 V < G/2$). However, this argument would not seem conclusive in the case of atoms of atomic number only slightly greater than 6, as it appears to be based only on order of magnitude considerations; also, there is no guarantee that G actually is less than $7eV.$; for example, Seitz (1940) gives $12eV.$ as a lower limit for excitation into the ionization levels of sodium chloride. Moreover, for all atomic numbers, there should be outer traps in the ionized atom, as discussed in § 2. The probability of capture into a trap increases with decreasing atomic number, and one might suppose it to be considerable for $Z < 6$, when F.T. also predict traps in the neutral atom, and fairly small otherwise.

The time spent in a trap can be assigned an upper limit of about 10^{-6} sec., on account of radiation loss, as given by eqn. (9). The figure may be consulted for the time spent in a trap if Auger loss alone is effective.

* Strictly, the curves apply only to sulphur, as L depends also on p_0 and the ionic radius of the element, but the curves for other elements are not expected to be much displaced.

Times less than 10^{-8} sec. are predicted for meson orbits of all radii almost up to the atomic radius of sulphur, provided G is smaller than 8 eV. Considering the crudity of the approximation, and the uncertainty of G , perhaps the best overall prediction one can make is that the time spent in a trap in the outer regions of the atom could well be much longer than the F.T. time of approach of 10^{-13} sec.; and the possibility exists that in insulators with strongly bound electrons, there may be traps of times comparable with the decay times, approximately 10^{-6} sec. and 10^{-8} sec. of the μ and π mesons respectively. Capture into such traps would, however, be rather improbable in heavy atoms.

The experimental evidence mentioned in §1 shows that, in fact, no appreciable number of mesons is captured into traps with times comparable with 10^{-8} sec., at any rate for $Z \geq 6$. For the purpose of further predictions (*e. g.* dealing with $900 m_e$ mesons in §4), we must assume that the values of G and the effective maximum orbital radius a are such as to be compatible with this result.

§4. MESONS OF MASS $900 m_e$.

For metals the time of approach of a $900 m_e$ mass meson has been estimated, applying the methods of F.T. The range of energy dealt with is downwards from +9000 eV., when the meson velocity is approximately equal to that of the valence electrons. The time of approach in a given metal is almost simply proportional to the meson mass, since it is proportional to the characteristic time t_0 [eqn. (1)]. In fact, in graphite the time of approach for a $900 m_e$ meson comes out to be 4.6×10^{-13} sec. (*cf.* 9.2×10^{-14} sec. for $200 m_e$ meson); and in iron to be 3.2×10^{-13} sec. (*cf.* 6.1×10^{-14} sec. for $200 m_e$ meson).

In insulators, if traps are ignored, an average rate of loss at each meson energy can in principle be obtained by F.T.'s methods. This would require calculation of the mean kinetic energy \bar{T} (§2), involving considerable computation when the absolute value of the meson energy is small. The increase of time in insulators as compared with metals is greater for heavy mesons than for light, because owing to the smaller meson velocity there are larger regions of the atom where loss cannot occur, on account of eqn. (2). Rough estimates show that the time of approach in insulators with $Z \geq 6$ is unlikely to be more than double that in a metal; *i. e.*, the time of approach will lie between 10^{-13} and 10^{-12} sec.

With regard to the existence of traps, even F.T.'s arguments lead to traps in atoms of atomic number up to 10, if $G=7$ eV., or higher if G is greater. In addition, there will be outer stable orbits in all elements if the atom is ionized, and these are even more likely to constitute traps for heavy mesons than for light.

The probability of capture into a trap can be supposed quite appreciable in light atoms and rather small in heavy ones, a dividing line being drawn at about $Z=10$, where F.T.'s theory would place the appearance of traps.

The upper limit for the time spent in a trap, due to radiation loss, is of the order of 10^{-5} sec. for these mesons. The trapping time, considering only loss to electrons, as shown in the figure, is generally much longer than the "average" time for untrapped mesons, viz., 10^{-12} , 10^{-13} sec., even for G as low as 4eV. If G is larger than 6eV., the time predicted is much greater even than the radiative time 10^{-5} sec., for a wide range of orbital radii; and certainly there is a wide range of combinations of G and a which are compatible with the requirement that the time for a $200 m_e$ meson should be less than 10^{-8} sec., and yet which yield a time much greater than 10^{-5} sec. for a $900 m_e$ meson.

Considering the crudity of the calculations, then, the overall conclusion is that in insulators many mesons will have approach times of 10^{-12} – 10^{-13} sec.; but a number of mesons, greater in light media than in heavy, may be trapped for considerably longer times, which may quite well be as much as 10^{-5} sec.

ACKNOWLEDGEMENT.

I should like to express my indebtedness to Professor H. Fröhlich for his suggestion of this work, and for his continued interest. My thanks are also due to Professor N. F. Mott for helpful discussions.

APPENDIX I.

RATE OF CHANGE OF ENERGY OF AN ELECTRON IN THE FIELD OF AN ATOM AND A ROTATING POINT CHARGE.

Consider an electron in the field of an atom (or ion) and a point charge (meson) $-e$, rotating about the nucleus with small uniform angular velocity ω , in a circular orbit of radius a . The potential due to the atom (not counting the field of the meson) is centrally symmetric, the total charge of the nucleus and the electrons other than the one under consideration being greater than $+e$.

Take polar coordinates (r, θ, ϕ) for the electron, with the origin at the nucleus. The principal axis is taken normal to the plane of rotation of the meson, and the plane from which ϕ is measured is chosen so that the meson is crossing it at time $t=0$. If $H(r, \theta, \phi)$ is the Hamiltonian of the electron at $t=0$, and includes the meson-electron interaction potential, the Schrödinger equation is

$$H(r, \theta, \phi - \omega t) \Psi(r, \theta, \phi, t) = -\frac{\hbar}{i} \frac{\partial \Psi(r, \theta, \phi, t)}{\partial t} \quad (10)$$

Transform to coordinates rotating with the meson, by putting

$$\phi - \omega t = \Phi; \quad \Psi(r, \theta, \phi, t) = \Omega(r, \theta, \Phi, t) \quad (11)$$

Then (10) becomes

$$[H(r, \theta, \Phi) - \omega M_z] \Omega(r, \theta, \Phi, t) = -\frac{\hbar}{i} \frac{\partial \Omega(r, \theta, \Phi, t)}{\partial t} \quad (12)$$

where M_z is the angular momentum operator

$$M_z = +\frac{\hbar}{i} \frac{\partial}{\partial \Phi}.$$

Now expand $\Omega(r, \theta, \Phi, t)$ in terms of "adiabatic" wave functions, following a procedure by Mott (1931) :—

$$\Omega(r, \theta, \Phi, t) = \sum_p a_p(t) \psi_p(r, \theta, \Phi) + \sum_K \int_0^\infty a_{K, W}(t) \psi_{K, W}(r, \theta, \Phi) dW, \quad (13)$$

where the discrete-state wave functions ψ_p satisfy

$$H(r, \theta, \Phi) \psi_p(r, \theta, \Phi) = W_p \psi_p(r, \theta, \Phi); \quad (14)$$

and the continuous-spectrum wave functions $\psi_{K, W}$ satisfy a similar equation, in which W_p is replaced by W . The wave functions $\psi_{K, W}$ are so normalized that

$$\int \psi_{K', W'}^* \psi_{K, W} d\tau = \delta_{K, K'} \delta(W - W'). \quad (15)$$

Substitution of (13) in (12) leads to the following equations for the a 's :—

$$\left. \begin{aligned} \dot{a}_p &= \frac{i}{\hbar} \left\{ -W_p a_p + \omega \left[\sum_q a_q(p | M_z | q) \right. \right. \\ &\quad \left. \left. + \sum_{K'} \int_0^\infty a_{K', W'}(p | M_z | K', W') dW' \right] \right\}, \\ \dot{a}_{K, W} &= \frac{i}{\hbar} \left\{ -W a_{K, W} + \omega \left[\sum_q a_q(K, W | M_z | q) \right. \right. \\ &\quad \left. \left. + \sum_{K'} \int_0^\infty a_{K', W'}(K, W | M_z | K', W') dW' \right] \right\}. \end{aligned} \right\} \quad (16)$$

Now we make the following simplifications :—

(i) The electron is initially in a non-degenerate adiabatic level, described by $p=0$. (The case of degeneracy is dealt with later.)

(ii) Where degeneracy in excited states occurs, the wave functions ψ are especially chosen so that the matrix elements of M_z connecting states of the same energy are zero.

(iii) On the right-hand side of (16), all terms containing matrix elements are ignored unless either

(a) one of the states involved is the initial state $p=0$, or

(b) the element is diagonal. In the continuous spectrum this means a δ function,

$$(K, W | M_z | K, W') = M_z(K, W) \delta(W - W'). \quad (17)$$

This leads to the solution for the continuous spectrum coefficients

$$a_{K, W} = \frac{i\omega t}{\hbar} (K, W | M_z | 0) \exp \left[\frac{-it}{2\hbar} (W_0 + W - \omega(0 | M_z | 0) - \omega M_z(K, W)) \right] \\ \times \frac{\sin \frac{t}{2\hbar} (W - W_0 - \omega M_z(K, W) + \omega(0 | M_z | 0))}{\frac{t}{2\hbar} (W - W_0 - \omega M_z(K, W) + \omega(0 | M_z | 0))} \quad (18)$$

Resonance transitions occur when the denominator vanishes. The energy transfer can now be calculated from the change in the expectation value of the energy $\bar{H}(t)$

$$\Delta H = H(t) - H(0) = \sum_p (W_p - W_0) |a_p(t)|^2 + \sum_K \int_0^\infty (W - W_0) |a_{K, W}(t)|^2 dW. \quad (19)$$

The contributions to this from transitions into the discrete spectrum will be neglected. By use of (18), ΔH can be approximately calculated by familiar methods, the result being

$$-\frac{dW}{dt} = \frac{\Delta \bar{H}}{t} = \frac{2\pi\omega}{\hbar} \sum_K \frac{|(K, W_K | [M_z, V] | 0)|^2}{(M_z(K, W_K) - (0 | M_z | 0)) \left(1 - \omega \frac{dM_z(K, W)}{dW} \Big|_{W=W_K} \right)} \quad (20)$$

Here W_K is the resonance energy with which an electron is ejected into the K state, and is given by

$$W_K = W_0 + \omega[M_z(K, W_K) - (0 | M_z | 0)]. \quad (21)$$

The sum is to be taken over such K as satisfy

$$W_K \geq 0.$$

In (20), $[M_z, V']$ is the commutator

$$M_z V' - V' M_z = \frac{\hbar}{i} \frac{\partial V'(\vec{r}, \theta, \Phi)}{\partial \Phi}, \quad (22)$$

(where V' is the interaction potential between meson and electron) and it has been introduced by means of the relation

$$(K, W | M_z | 0) = \frac{-(K, W | [M_z, V'] | 0)}{W - W_0}. \quad (23)$$

We may note the modification necessary if the initial level is degenerate, or approximately so. If all ways of occupying the degenerate initial states are equally probable, then the energy losses for the individual states should be calculated according to (20), then divided by the number of states and added.

For the actual evaluation of the rate of energy loss, further simplifications are made. It is assumed that the angular dependence of the adiabatic wave functions is approximately spherical-harmonic, the corresponding quantum numbers being initially (l_0, m_0) and finally (L, M) . Then

$$(K, W | [M_z, V'] | 0) = \hbar(M - m_0)(K, W | V' | 0), \quad (24)$$

and (20) becomes

$$-\frac{dW}{dt} = 2\pi\omega \sum_{\substack{L \geq M \\ M \geq M_0 + m_0}} (M - m_0) | (L, M, W_M | V' | 0) |^2, \quad (25)$$

where

$$M_0 = \frac{|W_0|}{\hbar\omega}, \quad (26)$$

and

$$W_M = \hbar\omega(M - M_0 - m_0). \quad (27)$$

For the purpose of further approximations, M_0 will be assumed large, and generally an upper limit for $-dW/dt$ will be sought.

To evaluate the matrix elements in (25), the radial parts of the wave functions are chosen as follows:—

(a) Initial state. A Slater-type wave function is taken (Slater, 1930)

$$\rho_0(r) = \frac{(2\kappa)^{n+\frac{1}{2}}}{((2n)!)^{\frac{1}{2}}} r^{n-1} e^{-\kappa r}, \quad (28)$$

where n and κ are appropriate constants.

(b) Final state. The wave-function is that for an electron in the field of a point charge $Z'e$.

$$\begin{aligned} \rho_{L, M, W_M} = & \frac{1}{\hbar} \left\{ \frac{2mk_M}{\pi} \left(L - \frac{iZ'}{k_M a_H} \right)! \left(L + \frac{iZ'}{k_M a_H} \right)! \right\}^{\frac{1}{2}} \cdot \frac{(2k_M r)^L}{(2L+1)!} \\ & \times \exp \left(\frac{\pi Z'}{2k_M a_H} + ik_M r \right) \times F \left(L+1 - \frac{iZ'}{k_M a_H}, 2L+2, -2ik_M r \right), \end{aligned} \quad (29)$$

where

$$k_M = \left(\frac{2m_e W_M}{\hbar^2} \right)^{\frac{1}{2}}. \quad (30)$$

For simplification this is replaced by

$$\rho_{L, M, W_M} = \frac{1}{\hbar} \left(\frac{2m_e k_M}{\pi} \right)^{\frac{1}{2}} \frac{L!}{(2L+1)!} \cdot \exp \left(\frac{\pi Z'}{2k_M a_H} \right) \cdot (2k_M r)^L, \quad (31)$$

which for large enough M should yield a good approximation to the

APPENDIX II.

ESTIMATION OF VARIABLES.

The governing factor L [eqns. (5)–(7)] depends on the variables G , a , M_0 , Z' and l_0 . The last of these is not very important and has been given the representative value 1. As explained in § 2, Z' has been taken as 1. It is desired to find how L depends on G and a as parameters, and so no particular choice of these is necessary. It remains to calculate M_0 (eqn. (4)), which will depend on G and a , as well as on the mass of the meson and the features of the electron cloud. The angular velocity ω must first be found. The atom is assumed to have lost one electron, so that if the meson were at the periphery of the resulting ion ($a=a_{\text{ion}}$), it would be moving in the Coulomb field of charge $+e$. We should then have

$$\omega_{\text{periphery}} = \left(\frac{e^2}{\mu a_{\text{ion}}^3} \right)^{\frac{1}{2}} \quad \dots \quad (36)$$

If the field remained Coulombian further in, ω would vary as $a^{-3/2}$; but in fact it will vary more rapidly. The rate of variation is limited by the condition for stable orbits, viz., that ω decrease with a less rapidly than a^{-2} . In order to obtain an upper limit to the rate of energy transfer, we assume the border-line condition ($\omega \propto a^{-2}$), which in conjunction with (36) leads to

$$\hbar\omega = 2W_{\text{H}} \cdot \frac{a_{\text{H}}^2}{a^2} \cdot \left(\frac{m_e a_{\text{ion}}}{\mu a_{\text{H}}} \right)^{\frac{1}{2}} \quad \dots \quad (37)$$

Hence, and from (4), M_0 can be obtained as a function of G , a , and μ , provided a representative value of a_{ion} is chosen. The value actually used was that given by Slater's theory (1930) for an S^+ ion, viz., $3.50 a_{\text{H}}$.

To estimate the time spent in a trap in a typical case (200 m_e meson in an S^+ ion), $-dW/dt$ must first be calculated from (5), the value of A being obtained from (34) and (35). This requires values of the additional variables ν , n , and κ , on which, however, the rate does not depend sensitively. These are again obtained from Slater's theory (1930): the influence of the Coulomb field of the meson on the initial wave function is supposed the same as if the meson were at the centre of the atom, so that κ is the same as for the valence electrons in a neutral atom of phosphorus. The meson radius is assumed to be the atomic radius of sulphur, as given by Slater's theory. G is chosen more or less arbitrarily. The list of variables is then:—

$$G = 8\text{eV.}; \quad a = 1.65 a_{\text{H}}; \quad M_0 = 6.1; \quad Z' = 1; \quad l_0 = 1; \quad \nu = 3; \quad n = 3; \\ \kappa = 1.60/a_{\text{H}}.$$

The result is $s = 5.5$; $-dW/dt = 1$ unit W_{H} per 0.8×10^{-5} sec. This value of s means, according to (33), that most electrons go into a state with magnetic quantum number approximately 6.

To estimate the time spent in a trap, the variation of $-dW/dt$ with a should be found. The result in the typical case is

$$-\frac{dW}{dt} \propto \exp(-\gamma\delta a) \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

with $\gamma = 44/a_H$.

The total time in the trap of initial radius a is then obtained by integrating the loss over the path a with the rough result :—

$$\tau_{\text{trap.}} = 2W_H \left(\frac{a_H}{a} \right)^2 \frac{1}{\gamma a_H \left(-\frac{dW}{dt} \right)_a} \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

In the present case this is 10^{-7} sec.

From these data equation (8) can be derived and the curves of the figure can be plotted.

FERMI, E., TELLER, E., and WEISSKOPF, V., 1947, *Phys. Rev.*, **71**, 314.

F.T. FERMI, E., and TELLER, E., 1947, *Phys. Rev.*, **72**, 399.

FROHLICH, H., HUBY, R., KOŁODZIEJSKI, R., and ROSENBERG, R. L., 1948, *Nature, Lond.*, **162**, 450.

MOTT, N. F., 1931, *Proc. Camb. Phil. Soc.*, **27**, 553; 1949, *Proc. Phys. Soc. A.*, **62**, 136.

OCCHIALINI, G. P. S., and POWELL, C. F., 1948, *Nature, Lond.*, **162**, 168.

SEITZ, F., 1940, *The Modern Theory of Solids* (McGraw-Hill), p. 447.

SLATER, J. C., 1930, *Phys. Rev.*, **36**, 57.

TICHO, H. K., 1948, *Phys. Rev.*, **74**, 1337.

WHEELER, J. A., 1947, *Phys. Rev.*, **71**, 320.

LXVII. On Kramers' Formula for the Frequency Distribution of the Radiation Emitted by an Electron in a Parabolic Orbit.

By K. C. WESTFOLD,

Division of Radiophysics, Council for Scientific and Industrial Research,
Australia*.

[Received May 2, 1949.]

ABSTRACT.

It is shown that correction of Kramers' formula for emission during a free-free electron-ion encounter leads to a result which agrees with Gaunt's quantum-theory result for absorption by free electrons, as required by the correspondence principle.

The conditions under which Gaunt recovered Kramers' formula for emission in a parabolic orbit are not generally applicable.

The correct formula for the absorption per unit intensity by free electrons of unit density also agrees with Gaunt's formula if the factor $\ln(mg^2d/Ze^2)$ be replaced by $\ln(mg^2/2hf)$. There is the same relation between the classical and quantum formulæ for the spontaneous emissivity.

* Communicated by the Council for Scientific and Industrial Research of Australia.

IN the foregoing paper* it was shown that Kramers' formula for the frequency distribution of the radiation emitted by an electron during a near-parabolic orbit about a positive ion represents an invalid approximation. From this formula Kramers (1923) derived a formula for the effective cross-section for capture of an electron into one of the Bohr quantum orbits of an atom. Then he deduced the atomic absorption coefficient corresponding to the liberation of a bound electron by the incident radiation.

To take account of non-parabolic orbits he estimated a correction factor, but, apart from this, his formula for the atomic absorption coefficient has been adopted quite generally in astrophysical and ionospheric investigations. Further, his formula for emission in a parabolic orbit was recovered by Gaunt (1930) in an independent quantum-theory investigation. Gaunt also found the absorption per unit intensity of radiation by free electrons of unit density without any restriction on the type of orbit. Again, apart from a correction factor, he recovered that formula which follows from Kramers' invalid formula for the emission.

Some discussion of the significance of our result is evidently desirable. We shall show that the substitution of our formula for the emission also results in Kramers' formulæ for absorption by free electrons and the atomic absorption coefficient, apart from a correction factor, which corresponds significantly with that of Gaunt. Further, we find agreement between our formula and Gaunt's formula for absorption by free electrons whose azimuthal quantum number is large, as required by the correspondence principle. It will appear that the conditions taken by Gaunt to specify a parabolic orbit are too restrictive. Indeed, such conditions applied to the classical theory do result in Kramers' formula for the emission, but the corresponding orbit is hardly typical of all parabolic orbits.

Kramers derived his formula for the atomic absorption coefficient from the consideration that "a certain frequency interval in the radiation emitted on the classical theory corresponds with a process by which the electron is bound in a certain stationary state." If β_n be the effective cross-section for spontaneous capture of an electron of relative velocity g into the n th quantum state of an atom, the effective cross-section for absorption of radiation in the frequency range $(f, f+df)$ resulting in the liberation of an electron with velocity g from the n th quantum state is

$$\alpha_n = \frac{m^2 g^2 c^2}{2n(n+1)h^2 f^2} \beta_n, \quad \dots \dots \dots (1)$$

the atomic absorption coefficient. Here n is the principal quantum number of the state, h is Planck's constant, m the mass of an electron and c the free-space velocity of light. This relation is valid only under the condition $hf \gg kT$ when the stimulated emission is negligible. Kramers

* We shall refer to this paper as S.W.

applied the theory to X-ray spectra. The frequency interval for large values of n is approximately

$$\Delta_n f = \frac{4\pi^2 Z^2 e^4 m}{n^3 h^3}, \quad (2)$$

where e is the electronic charge and Z the atomic number. The units are Gaussian.

Then the correspondence is such that if q_n is the probability of the capture of an electron with the emission of a quantum hf ,

$$Q_f \Delta_n f = q_n hf, \quad (3)$$

where Q_f is the emission per unit frequency interval in an encounter. It follows that

$$\beta_n = 2\pi \int_0^{b_{\max}} q_n b \, db, \quad (4)$$

where b is the distance from the nucleus to the initial asymptote of the trajectory of the interacting electron.

On substituting the formula S.W. (3.27) for Q_f and integrating, we get

$$\beta_n = \frac{16\pi^2 Z^2 e^6}{3c^3 m^2 g^2 hf} \Delta_n f A_1(2) \quad (5)$$

where $A_1(2)$ is the same function as S.W. (3.7),

$$A_1(2) = \ln(1 + \nu_{01}^2), \quad (6)$$

where

$$\nu_{01} = \frac{mg^2 d}{Ze^2}, \quad (7)$$

d being the mean distance between pairs of interacting molecules of the ionized gas.

With Kramers' formula for Q_f we find

$$\beta_n = \frac{32\pi^2 Z^2 e^6}{3\sqrt{3}c^3 m^2 g^2 hf} \Delta_n f. \quad (5')$$

Thus we arrive at a correction factor for Kramers' formula for β_n ,

$$\gamma = \frac{\sqrt{3}}{2\pi} A_1(2). \quad (8)$$

Now in most physical situations it happens that $\nu_{01} \gg 1$. Hence under such conditions we may write

$$\gamma = \frac{\sqrt{3}}{\pi} \ln \nu_{01}. \quad (9)$$

We see from the relation (1) that γ is also the correction factor to be applied to Kramers' formula for the atomic absorption coefficient.

Using quantum theory, Gaunt paralleled the case of absorption in a parabolic orbit by considering the case of a free electron whose energy remains small and positive during an absorption process. Now, $a(g, b; f)$,

the time rate of absorption of radiation from a beam of f -radiation of unit intensity in a free-free electron-ion encounter, is related to the spontaneous emission $Q_f df$ by the formula

$$a(g, b; f) = \frac{Q_f df}{8\pi h f^3 df/c^2} = \frac{h_f}{4\pi} \beta_{12}/df, \quad \dots \quad (10)$$

where $\beta_{12} I_f$ is the probability of absorption of a quantum h_f during an encounter (S.W. (3.36)).

The parameters g, b specify the nature of the encounter. Then it is found that the value of a obtained by substituting Kramers' formula for Q_f in (10) is identical with Gaunt's result for electrons describing parabolic orbits. However, we can show that the discrepancy between Gaunt's result and ours is due to the difference between the two methods of approximation. Gaunt's parabolic case is specified by $E = \frac{1}{2} m g^2 \rightarrow 0$ while $h_f/E = 0(1)$ and he also assumes $gb = 0(E^{-1})$; our parabolic case is specified only by the sufficient condition $v_0 = m g^2 b / Z e^2 \ll 1$ and we have assumed* that $fb/g \ll 1$.

Under Gaunt's conditions $v_0 = 0(E^{\frac{1}{2}})$, which satisfies our condition, but other consequences ensue. If we put $z = \cot \theta/2$ in S.W. (3.24), approximations to J_{\parallel} and J_{\perp} for small v_0 are, by S.W. (3.16),

$$\left. \begin{aligned} J_{\parallel} &= \frac{2Ze^2}{mgb} \int_{-2/v_0}^{2/v_0} \exp \left\{ -\pi i \frac{fb}{g} v_0^2 (z + \frac{1}{3} z^3) \right\} \frac{z^2 - 1}{(z^2 + 1)^2} dz, \\ J_{\perp} &= \frac{2Ze^2}{mgb} \int_{-2/v_0}^{2/v_0} \exp \left\{ -\pi i \frac{fb}{g} v_0^2 (z + \frac{1}{3} z^3) \right\} \frac{2z}{(z^2 + 1)^2} dz. \end{aligned} \right\} \quad (11)$$

Under Gaunt's conditions the factor in the exponent, $(fb/g)v_0^2 = 0(1)$, while the limits $\pm 2/v_0 = 0(E^{\frac{1}{2}})$. As $E \rightarrow 0$ the limits become infinite and Kramers' result follows. Under our conditions, however, $(fb/g)v_0^2 \ll 1$ and the first approximation involves replacing the exponential term by unity, but the limits may not be replaced by $\pm \infty$. Gaunt's assumption that $gb = 0(E^{-1})$ as $E \rightarrow 0$ is evidently too restrictive a condition on parabolic orbits.

Gaunt also studied the case of absorption by free electrons under the sole condition that $f \rightarrow 0$. As the orbit is no longer restricted, this should correspond exactly with the conditions of our investigation.

He finds

$$a(g, b; f) = \frac{8\pi Z^2 e^6}{3ch^3 f^3} \left\{ \frac{k+1}{(k+1)^2 + n^2} + \frac{k}{k^2 + n^2} \right\} / (2k+1), \quad \dots \quad (12 a)$$

where

$$mgb = k\hbar/2\pi, \quad \dots \quad (13)$$

* In S.W. we have also made the assumption that the total energy Q radiated during an encounter is such that $Q \ll E$.

k being the azimuthal quantum number. Our result follows from S.W. (3.27) and (10). It may be written

$$a(g, b; f) = \frac{8\pi Z^2 e^6}{3ch^3 f^3} \cdot \frac{\hbar^2}{4\pi^2 m^2 g^2 b^2} \cdot \frac{v_0^2}{1+v_0^2} \cdot \dots \quad (12\ b)$$

According to the correspondence principle, (12 *a*) must agree with the classical result for large values of the quantum number k . Classically this corresponds to orbits showing small deflections, where v_0 is large. Under these conditions both (12 *a*) and (12 *b*) give approximately

$$a(g, b; f) = \frac{8\pi Z^2 e^6}{3ch^3 f^3} \cdot \frac{1}{k^2}, \quad \dots \quad (14)$$

as required by the correspondence principle. We have seen that Kramers' second formula for Q_f in such an orbit agrees with ours. It therefore agrees with Gaunt's result too. This contradicts a statement made by Gaunt where he considered Kramers' second formula.

The rate of absorption of energy from a beam of f -radiation of unit intensity by electrons having speed g and unit density is given by

$$a(g, f) = 2\pi \int_0^{b_{\max}} a(g, b; f) g b \, db. \quad \dots \quad (15)$$

Substituting (12 *b*) into (15) we find

$$a(g, f) = \frac{2Z^2 e^6}{3cm^2 ghf^3} A_1(2), \quad \dots \quad (16\ b)$$

which may be compared with Gaunt's result

$$a(g, f) = \frac{4Z^2 e^6}{3cm^2 ghf^3} \ln \left(\frac{E}{\hbar f} \right) \quad (\hbar f \ll E). \quad \dots \quad (16\ a)$$

If $v_{01} \gg 1$ we see that v_{01} corresponds with $E/\hbar f$, that is, Ze^2/d corresponds with $2\hbar f$. Using Kramers' parabolic formula for Q_f we find

$$a(g, f) = \frac{4\pi Z^2 e^6}{3\sqrt{3}cm^2 ghf^3} \cdot \dots \quad (16')$$

Gaunt's correction factor is therefore

$$\gamma_a = \frac{\sqrt{3}}{\pi} \ln \left(\frac{E}{\hbar f} \right) \quad \dots \quad (17\ a)$$

and ours is

$$\gamma_b = \frac{\sqrt{3}}{\pi} \ln v_{01}, \quad \dots \quad (17\ b)$$

the same as (9) for absorption by a bound electron.

Thus we have found that the correct approximation to the frequency distribution of radiation emitted spontaneously during a free-free electron-ion encounter, obtained from classical electromagnetic theory,

leads to a formula for absorption which is equivalent to that obtained from quantum theory under such conditions as are required by the correspondence principle. To obtain the quantum formulæ for the absorption per unit intensity by free electrons of unit density and the corresponding emissivity, the factor $A_1(2)$ in the classical formulæ must be replaced by $2\ln(E/hf)$. These formulæ may not be validly applied to a medium in which the proportion of multiple to binary encounters is not small. Where they are applicable the quantum formula resolves the arbitrariness of our selection of the minimum deflection which enables an encounter to be considered. It amounts to choosing $b_{\max} = Ze^2/2hf$ instead of d for the purpose of investigating the emissivity. In circumstances where $Ze^2/2hf \gg d$ we may expect the quantum formula to be misleading.

ACKNOWLEDGMENT.

The work described in this paper is part of the research programme of the Division of Radiophysics, Council for Scientific and Industrial Research, Australia. It has benefited from discussion with Mr. S. F. Smerd of this laboratory.

REFERENCES.

- KRAMERS, H. A., 1923, *Phil. Mag.*, **46**, 836.
GAUNT, J. A., 1930, *Phil. Trans. A*, **229**, 163.

LXVIII. *Interchange Interaction and Collective Electron Ferromagnetism.*

By E. P. WOHLFARTH, Ph.D.,
Department of Physics, Leeds University*.

[Received April 8, 1949.]

§ 1. INTRODUCTION.

THE aim in this paper is to give a discussion of the relation between electron interchange interaction and the ferromagnetic properties of metals. The problem of ferromagnetism has frequently been discussed from the point of view of the Heitler-London-Heisenberg approximation (see, for example, Sommerfeld and Bethe 1933), but the recent development of the collective electron treatment of ferromagnetism (Stoner 1938, 1939, 1948, Wohlfarth 1949 a) makes it desirable to consider how interchange interaction enters in the collective electron scheme. The whole problem is extremely complicated, and only preliminary consideration can be given here to a few of its aspects. After a brief survey of the general theoretical background in § 2, the problem is first discussed (§ 3) for perfectly free electrons in metals, this being the most convenient approach to the general collective electron scheme in its present formulation.

* Communicated by Professor E. C. Stoner.

The extension of the treatment to actual metals is discussed in §4, and experimental results are considered in the light of the general treatment in §5.

§2. THEORETICAL BACKGROUND.

While the Heitler-London-Heisenberg method of approximating to the electronic wave functions and energies of metals is based on that applicable to diatomic molecules, the most convenient approach to the metallic state in the collective electron scheme is through the treatment of many-electron atoms. With a view to applying this treatment to metals, Brillouin (1932, 1933, 1934) and others have made a suitable extension of Hartree's equations, taking for the initial wave function for the metal a determinant, the elements of which are one-electron wave functions of the Bloch type (1928). This wave function, in assuming which the exclusion principle is automatically satisfied, may be written in the form

$$\psi = \frac{1}{(N!)^{\frac{1}{2}}} \begin{vmatrix} \phi_1(\mathbf{r}_1, s_1) & . & . & . & . & \phi_1(\mathbf{r}_N, s_N) \\ . & & & & & . \\ . & & & & & . \\ . & & & & & . \\ . & & & & & . \\ . & & & & & . \\ \phi_N(\mathbf{r}_1, s_1) & . & . & . & . & \phi_N(\mathbf{r}_N, s_N) \end{vmatrix}, \quad . \quad . \quad . \quad (2.1)$$

where the element $\phi_n(\mathbf{r}_i, s_i)$ is the wave function for an electron in the state n and with spatial coordinate \mathbf{r}_i , measured from a fixed origin (the same for all electrons), and spin coordinate s_i . In the absence of strong spin-orbit coupling the ϕ are separable into spatial and spin functions

$$\phi_n(\mathbf{r}_i, s_i) = \psi_n(\mathbf{r}_i) \sigma_n(s_i), \quad . \quad . \quad . \quad . \quad . \quad (2.2)$$

and if the ψ are of the Bloch form,

$$\psi_n(\mathbf{r}_i) = \exp(i\mathbf{k}_n \cdot \mathbf{r}_i) u(\mathbf{k}_n, \mathbf{r}_i), \quad . \quad . \quad . \quad . \quad . \quad (2.3)$$

where u has the period of the lattice, \mathbf{k}_n being the electronic momentum, then mutual orthogonality is ensured, since the orthogonality conditions are obeyed by both the ψ and the σ .

The Hamiltonian operator for the assembly is

$$H = \sum_n \left\{ \left(-\frac{\hbar^2}{8\pi^2 m} \right) \nabla_n^2 + V_n \right\} + \frac{1}{2} \sum_m \sum_n \frac{e^2}{\mathbf{r}_{ij}}, \quad . \quad . \quad . \quad (2.4)$$

where V_n is the potential at \mathbf{r}_i of the nuclei and \mathbf{r}_{ij} the inter-electronic distance $|\mathbf{r}_i - \mathbf{r}_j|$. The total energy is then given by

$$U = \int \dots \int \Psi^* H \Psi d\tau(\dots \mathbf{r}_i \dots s_i \dots), \quad . \quad . \quad . \quad (2.5)$$

which, on inserting (2.1), gives

$$\begin{aligned}
 U = & \sum_n \left(-\frac{\hbar^2}{8\pi^2 m} \right) \int \phi_n^*(\mathbf{r}_i, s_i) \nabla_n^2 \phi_n(\mathbf{r}_i, s_i) d\tau(\mathbf{r}_i, s_i) \\
 & + \sum_n \int \phi_n^*(\mathbf{r}_i, s_i) V_n \phi_n(\mathbf{r}_i, s_i) d\tau(\mathbf{r}_i, s_i) \\
 & + \frac{1}{2} \sum_{m,n} \iint \frac{e^2}{r_{ij}} |\phi_n(\mathbf{r}_i, s_i)|^2 |\phi_m(\mathbf{r}_j, s_j)|^2 d\tau(\mathbf{r}_i, s_i; \mathbf{r}_j, s_j) \\
 & - \frac{1}{2} \sum_{m,n} \iint \frac{e^2}{r_{ij}} \phi_n^*(\mathbf{r}_i, s_i) \phi_m^*(\mathbf{r}_j, s_j) \phi_m(\mathbf{r}_i, s_i) \phi_n(\mathbf{r}_j, s_j) d\tau(\mathbf{r}_i, s_i; \mathbf{r}_j, s_j). \quad (2.6)
 \end{aligned}$$

The first term in (2.6) represents the kinetic energy, E , of the electrons ; it may be simplified by using (2.2) and the orthogonality conditions for the spin functions

$$\iint \sigma_n(s_i) \sigma_n(s_j) d\tau(s_i, s_j) = \delta_{ij}; \quad . \quad . \quad . \quad . \quad (2.7)$$

giving

$$E = \sum_n \left(-\frac{\hbar^2}{8\pi^2 m} \right) \int \psi_n^*(\mathbf{r}_i) \nabla_n^2 \psi_n(\mathbf{r}_i) d\tau(\mathbf{r}_i). \quad . \quad . \quad . \quad . \quad (2.8)$$

The second and third terms in (2.6) represent the Coulomb energy due to the nuclei and electron charge cloud respectively. These terms are of obvious importance in calculating cohesive energies, for example (cf. Mott and Jones 1936), but they are independent of the magnetization (see below) and may conveniently be omitted in discussing ferromagnetism, by an appropriate re-definition of the energy zero.

The final term in (2.6) represents the interchange interaction energy, J ; its inclusion in U is a direct consequence of the use of the determinantal wave function (2.1), *i. e.* of the exclusion principle, and is thus compatible with the subsequent application of Fermi-Dirac statistics to determine the distribution of the electrons among the different \mathbf{k} states. The expression for J may be simplified by the use of (2.7), giving

$$J = \sum_{m,n} J_{mn},$$

where

$$J_{mn} = -\frac{1}{2} \iint \frac{e^2}{r_{ij}} \psi_n^*(\mathbf{r}_i) \psi_m^*(\mathbf{r}_j) \psi_m(\mathbf{r}_i) \psi_n(\mathbf{r}_j) d\tau(\mathbf{r}_i, \mathbf{r}_j), \quad . \quad . \quad (2.9)$$

if the spins i and j are parallel, but giving $J_{mn} = 0$ if they are anti-parallel. In this representation, therefore, which corresponds to a particular approximation, the interchange interaction forces are effective only between electrons with parallel spins ; if N_1, N_2 are the numbers of electrons having the two directions of spin respectively, then

$$J = \sum_{m=1}^{N_1} \sum_{n=1}^{N_1} J_{mn} + \sum_{m=1}^{N_2} \sum_{n=1}^{N_2} J_{mn} \quad (m \neq n), \quad . \quad . \quad . \quad (2.10)$$

with J_{mn} given by (2.9).

The interchange interaction forces have frequently been termed "correlation forces", since the simplest way of picturing them is by considering their action in correlating the positions of the individual electrons (*cf.* Slater 1934 a, Seitz 1940, *et al.*). In this terminology the energy calculated using the initial wave function (2.1) contains correlation terms for parallel spins only. It is plausible to suppose that there is some correlation for electrons with anti-parallel spins also, but the above discussion shows that the corresponding energy term can only be calculated by using a more rigorous initial wave function. This problem has so far been treated only by Wigner (1934; *cf.* also Seitz 1940), who considers the inter-electronic distances explicitly and includes in his initial wave function terms of the form $\psi_n(\mathbf{r}_i; \mathbf{r}'_1, \mathbf{r}'_2, \dots \mathbf{r}'_i \dots)$, where \mathbf{r}_i refers to the position of an electron with spin to the right, say, and \mathbf{r}'_i with spin to the left. The perturbation calculation is here much more difficult and was carried out approximately for free electrons only. Wigner assumes that the difference in energy calculated by using the more complicated wave function and that derived from (2.1) is due entirely to anti-parallel spin correlation, that is, that the parallel spin correlation energy remains unchanged by the additional terms in the initial wave function. This assumption is open to criticism, and, furthermore, using a more complicated wave function than (2.1), it is in general not possible to satisfy the exclusion principle, so that the relationship between the quantum mechanical and statistical parts of the treatment becomes less obvious.

§ 3. APPLICATION TO FREE ELECTRONS IN METALS.

The evaluation of the relevant energy terms is most easily carried out for perfectly free electrons. While the results of such calculations are not immediately applicable to actual ferromagnetics, which are discussed in §§ 4 and 5, they do yield valuable qualitative information about points of interest in connection with the statistical treatment.

For perfectly free electrons, $u(\mathbf{k}_n, \mathbf{r}_i)$ in (2.3) is constant and the normalized one-electron wave functions take the form

$$\psi_n(\mathbf{r}_i) = \Omega^{-1/2} \exp(i\mathbf{k}_n \cdot \mathbf{r}_i), \quad (3.1)$$

where Ω is the volume of the metal. The kinetic energy is given by (2.8), leading to

$$E = \sum_n E_n = \sum_n (\hbar^2 k_n^2 / 8\pi^2 m). \quad (3.2)$$

In this case the Fermi surface in momentum space is spherical, with radius k_0 , which, in the absence of spontaneous magnetization, is given by $k_0 = (3\pi^2 N / \Omega)^{1/3}$, while, if magnetization is complete, *i. e.* all the spins are aligned, $k_0 = (6\pi^2 N / \Omega)^{1/3}$. In the intermediate case, if N_1 spins are directed to the left, say, and N_2 to the right, there is a net spontaneous magnetization of amount $N\mu\zeta$, where

$$\zeta = (N_2 - N_1) / N = (N_2 - N_1) / (N_2 + N_1).$$

The kinetic energy at absolute zero is easily shown from (3.2) to be

$$E(\zeta) = \frac{3}{10} N \epsilon_0 f_{5/3}(\zeta), \quad \dots \dots \dots (3.3)$$

where N is the total number of electrons in volume Ω , ϵ_0 the zero point energy in the absence of spontaneous magnetization, given by

$$\epsilon_0 = (\hbar^2/8m)(3N/\pi\Omega)^{2/3},$$

and where

$$f_\nu(x) = \{(1+x)^\nu + (1-x)^\nu - 2\}.$$

Here, as elsewhere, energies are measured relative to the completely demagnetized state, so that, for example, $E(0)=0$ (cf. also Stoner 1939, eq. (4.11) and (4.12)).

The interchange interaction term (2.9) for perfectly free electrons takes the form

$$J_{mn} = -\frac{1}{2}(e/\Omega)^2 \iint_{\Omega} \frac{1}{r_{ij}} \exp \{i(\mathbf{k}_m - \mathbf{k}_n) \cdot (\mathbf{r}_i - \mathbf{r}_j)\} d\tau_i d\tau_j.$$

This integral was first evaluated by Bloch (1929; cf. also Sommerfeld and Bethe 1933), and by considering the two sets of spins and carrying out the summation over all the occupied states (cf. (2.10)), Bloch's result may be put in the form

$$\text{where } \left. \begin{aligned} J(\zeta) &= -\frac{3}{8} N \epsilon_j f_{4/3}(\zeta), \\ \epsilon_j &= e^2 (3N/\pi\Omega)^{1/3}, \end{aligned} \right\} \dots \dots \dots (3.4)$$

the energy being again measured relative to the demagnetized state. The magnetization dependent part of the total energy at absolute zero is thus given by $U(\zeta) = E(\zeta) + J(\zeta)$, giving, for free electrons,

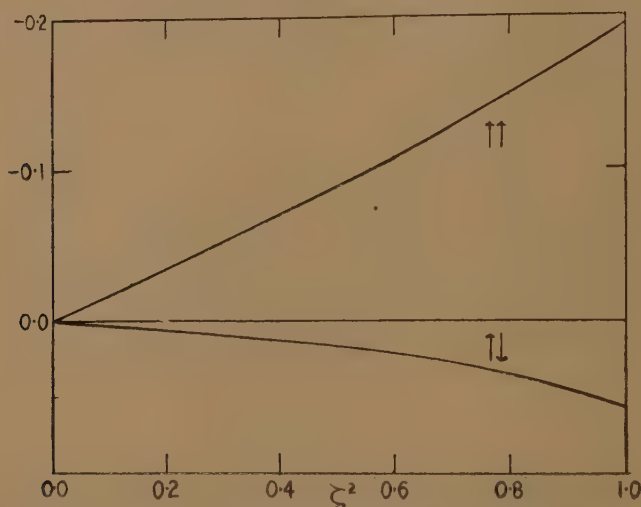
$$U(\zeta)/N\epsilon_0 = \frac{3}{10} \left\{ f_{5/3}(\zeta) - \frac{5}{4} \frac{\epsilon_j}{\epsilon_0} f_{4/3}(\zeta) \right\}. \quad \dots \dots \dots (3.5)$$

On differentiating (3.5) it is found that $U(\zeta)$ has a minimum for $\zeta=0$, so that there is no net spontaneous magnetization, for $\epsilon_j/\epsilon_0 < 2$, but that for $\epsilon_j/\epsilon_0 \geq 2$ it has a maximum at $\zeta=0$ and attains its lowest value for $\zeta=1$. For $\epsilon_j/\epsilon_0 \geq 2$, therefore, all the spins are parallel, *i. e.* saturation is complete at absolute zero. For reasons given below, the transition to complete saturation is probably less abrupt than is deduced from (3.5).

As pointed out above, the total energy must be supplemented by a term giving the interchange interaction (or correlation) energy for electrons with anti-parallel spins. The results of Wigner's calculation which, as stated in § 2, gives at best only a rough approximation, apply to an assembly of perfectly free electrons with zero spontaneous magnetization, and the extension to a state with magnetization ζ would require further calculation. An approximate expression has been given by Seitz (1940, p. 600), but this is reliable only for ζ small. It is obvious, however, that the Wigner term must vanish for $\zeta=1$.

In fig. 1 the lower curve represents schematically the variation of the Wigner term with the square of the relative magnetization, ζ^2 , measured relative to the energy of the demagnetized state. The upper curve gives the variation of the parallel spin term $J(\zeta)/N\epsilon_j = -\frac{3}{8}f_{4/3}(\zeta)$ with ζ^2 (cf. (3.4)), measured from the same zero. The total interchange interaction energy is the sum of these two terms, and the form of its variation with ζ^2 depends on their relative magnitudes. If the Wigner term is relatively unimportant, the slope of the total interchange interaction, ζ^2 curve will increase with ζ^2 , while if it is fairly large the slope will decrease. The intermediate

Fig. 1.



Variation of interchange interaction energy terms with square of the magnetization.

Lower curve: Wigner term for anti-parallel spins (schematic).

Upper curve: Bloch term for parallel spins, cf. (3.4). ζ , relative magnetization at absolute zero.

case, in which the total interchange interaction energy is strictly proportional to the square of the magnetization, is the one considered by Stoner (1938), who writes the total interchange interaction term in the form $-\frac{1}{2}Nk\theta'\zeta^2$, with θ' a constant coefficient. The total energy at absolute zero is in this case given by

$$U(\zeta)/N\epsilon_0 = \frac{3}{10}f_{5/3}(\zeta) - \frac{1}{2}(k\theta'/\epsilon_0)\zeta^2, \quad \dots \quad (3.6)$$

and the equilibrium state by $dU(\zeta)/d\zeta = 0$, i. e.

$$(1+\zeta)^{2/3} - (1-\zeta)^{2/3} = 2(k\theta'/\epsilon_0)\zeta, \quad \dots \quad (3.7)$$

which is the equation derived previously (Stoner 1938, eq. (5.10)). In this case, as shown by (3.7), the energy minima lie between 0 and 1 for

$$\frac{3}{2} < \frac{k\theta'}{\epsilon_0} < 2^{-1/3},$$

so that, for this range of the magnetization dependent part of the interchange interaction energy, saturation at absolute zero is not complete. Although the expressions (3.5) and (3.6) are very similar (*cf.* fig. 1) the consequences to which they lead are strikingly different, since, as already pointed out, using (3.5) saturation is suddenly attained if a particular value of the interchange interaction energy is exceeded, while, with (3.6), the transition is less abrupt.

Stoner's statistical treatment, which is based essentially on (3.6) as the appropriate expression for the energy, has been applied with considerable success to the interpretation of several magnetic and thermal properties of particular metals and alloys (Wohlfarth 1948, 1949a). Before considering this work in the light of the present discussion, it is necessary to see how the formulæ derived for free electrons have to be modified when applied to electrons in actual metals.

§ 4. APPLICATION TO ACTUAL METALS.

The discussion in this section will be confined to parallel spin interaction energies, since it is not possible at present to deal with the anti-parallel spin terms in more than a qualitative way. It should eventually be possible, however, to assess the approximation involved in neglecting these uncertain terms by comparing the results of calculations of the parallel spin interaction energies with those derived experimentally (*cf.* § 5).

General considerations.

The formulæ given in the preceding section may be applied approximately to metals like sodium and copper, the outermost *s* electrons of which are very nearly free; they also apply to the *s* electrons of transition metals like nickel and palladium. It was shown in conjunction with relation (3.5) that the "condition for ferromagnetism" may here be written as $\epsilon_j/\epsilon_0 > 2$, which, inserting (3.3) and (3.4) and numerical values, gives

$$\epsilon_j/\epsilon_0 = 0.4604(A/q\rho)^{1/3} > 2, \quad . \quad . \quad . \quad . \quad . \quad (4.1)$$

where *A* is the atomic weight, ρ the density and *q* the number of electrons per atom. For the 4*s* electrons of nickel, $q \doteq 0.6$, and hence $\epsilon_j/\epsilon_0 \doteq 1.03$, showing that the ferromagnetism of nickel cannot be due to the *s* electrons. This conclusion, which is, of course, not new, can be stated in a slightly different form using (3.4). For $\zeta \ll 1$, $J(\zeta) = -\frac{1}{6}N\epsilon_j\zeta^2$, which, identified with Stoner's relation for $J(\zeta)$, gives $k\theta' = \epsilon_j/3 = (e^2/3)(3N/\pi\Omega)^{1/3}$. For the 4*s* electrons in nickel, therefore, $\theta' \doteq 2.1 \times 10^3^\circ \text{K.}$, which is roughly of the order of the Curie temperature which nickel would have if its ferromagnetism were due to the *s* electrons. A value of the same order of magnitude was derived by Stoner (1930) from a rather different standpoint, and similar conclusions were reached, that the ferromagnetism of nickel cannot be ascribed to its "conduction" electrons.

As is well known, the ferromagnetic properties of metals and alloys may be ascribed to the "holes" in their d bands, and, since these are to a large extent equivalent to electrons in otherwise empty bands, it is permissible to consider interchange interaction effects between holes in the same way as for electrons (for fuller discussion, see Stoner's papers). The principal distinction between the d and s bands of the transition metals arises from the very much higher density of states in the d band (Mott 1935, Slater 1936). For nickel, for example, the number of holes in the d band and the number of electrons in the s band are equal, but whereas the occupied bandwidth for the s electrons is roughly that calculated by the free electron relation, the results of low temperature specific heat and other measurements indicate that the unoccupied energy width of the d band of nickel is about 30 times less. Since, in nickel, the energy distribution of states near the Fermi limit in the d band may, with fair approximation, be taken to be of the same form as that for free electrons, it is convenient in this case to consider the holes as "quasi-free", with an effective mass, m^* , of about $30 m$, *i. e.* to express the unoccupied bandwidth in the form

$$\epsilon_0 = (\hbar^2/8m^*)(3N/\pi\Omega)^{2/3}. \quad (4.2)$$

It is this particular distribution which Stoner considers in detail in his statistical treatment.

It is very much more difficult to determine the influence of the large effective mass of the d electrons on their interchange interaction energy. It is obvious that it must be influenced to some degree, since otherwise the ratio ϵ_j/ϵ_0 would become impossibly large, attaining the value $1.03 m^*/m \doteq 30$ for nickel (*cf.* (4.1)). The possibility, which lays no claim to any theoretical rigour, has been considered of describing the holes in the d band by wave functions of the form

$$\psi_n(\mathbf{r}_i) = \Omega^{-1/2} \exp \{i(m^*/m)^{1/2} \mathbf{k}_n \cdot \mathbf{r}_i\}. \quad (4.3)$$

This leads to the correct expression for ϵ_0 , as in (4.2), and, by adaptation of Bloch's method, gives the value of the interchange interaction energy as in (3.4), but with

$$\epsilon_j = (me^2/m^*)(3N/\pi\Omega)^{1/3}. \quad (4.4)$$

To this approximation, therefore, ϵ_j/ϵ_0 has the same value as for perfectly free electrons (*cf.* above) and ferromagnetism is again precluded for nickel. On the other hand, the value of θ' is now reduced in the ratio (m^*/m) , giving, for nickel, $\theta' \doteq 700^\circ \text{K.}$, which is at least of the correct order of magnitude, both of the value deducible from the experimental results and of the Curie temperature.

The foregoing discussion shows clearly what the essential difficulties are: The interchange interaction energy for free electrons is too small to lead to ferromagnetism; the extension to "quasi-free holes" in the manner just described leads to no improvement in that respect. The indications are, therefore, that for significant values of the interchange interaction energy of a particular metal to be obtained, it is necessary

to consider in detail the actual wave functions of the metal. This would, in the first place, differentiate between d and s electrons, leading, in certain cases, to ferromagnetism for the former but not for the latter. It is, however, necessary to attain a much higher degree of precision, as is exemplified by nickel and palladium. The experimental results indicate (Wohlfarth 1948) that the electronic band structures of these two metals are strikingly similar, and that, in addition, the values of the interchange interaction energy are very close (values derived from experimental results, the magnetic properties of nickel-palladium alloys, give $\theta' \doteq 1.2 \times 10^3$ °K. for nickel and 1.0×10^3 °K. for palladium, a difference of only about 0.02 eV.). To attain the high degree of precision necessary to differentiate between two metals like nickel and palladium, calculations of considerable complexity have to be carried out, and in this paper only a very brief discussion can be given of some of the difficulties involved, leaving it to future work to consider in detail the procedure to be adopted for the discussion of particular metals and alloys.

Derivation of Interchange Interaction Energy by Approximational Methods.

There are at present three well known approximational methods for the description of the behaviour of electrons in metals in the collective electron scheme. These have been used with considerable success in the calculation of energy bands (see, for example, Mott and Jones 1936, or Seitz 1940), but have not so far been discussed in detail with a view to calculating interchange interaction energies.

(a) Approximation of nearly free electrons (Peierls 1930, *cf.* also Sommerfeld and Bethe 1933). Although this method is not likely to be accurate for transition metals, it does indicate some of the main difficulties, particularly those arising from the Brillouin zone structure, which may be of particular importance for ferromagnetic alloys.

The one-electron wave functions (2.3) may be written in the form of a Fourier series

$$\psi_n(\mathbf{r}_i) = \exp(i\mathbf{k}_n \cdot \mathbf{r}_i) \sum_{\mathbf{m}} c_{\mathbf{m}}(\mathbf{k}_n) \exp(2\pi i \mathbf{m} \cdot \mathbf{r}_i/a), \quad \dots \quad (4.4)$$

and the potential due to the lattice (in this approximation everywhere small)

$$V = \sum_{\mathbf{m}} V_{\mathbf{m}} \exp(2\pi i \mathbf{m} \cdot \mathbf{r}_i/a). \quad \dots \quad (4.5)$$

The perturbation calculation of energy shows that all the $c_{\mathbf{m}}$ are small compared with c_0 , unless the Bragg reflection condition

$$\mathbf{m} \cdot \mathbf{k}_n = \pi |\mathbf{m}|^2/a \quad \dots \quad (4.6)$$

is satisfied, indicating the presence of a Brillouin zone boundary, near which (4.4) may be written

$$\psi_n(\mathbf{r}_i) = \exp(i\mathbf{k}_n \cdot \mathbf{r}_i) \{c_0 + c_{\mathbf{m}} \exp(2\pi i \mathbf{m} \cdot \mathbf{r}_i/a)\}, \quad \dots \quad (4.7)$$

with the normalization condition

$$c_0^2 + c_{\mathbf{m}}^2 = \Omega^{-1}.$$

The coefficients c_0 and c_m are complicated functions of V_m and \mathbf{k}_n , and are such that $|c_m/c_0|$ is small if (4.6) is not satisfied, but $|c_m/c_0| \rightarrow 1$ in the neighbourhood of a zone boundary.

Considering now two electrons with wave functions $\psi_n(\mathbf{r}_i)$, $\psi'_n(\mathbf{r}'_i)$, a straightforward extension of the method used in evaluating (2.9) for free electrons gives, with $\psi_n(\mathbf{r}_i)$ given by (4.7),

$$\begin{aligned} J_{nn'} &= J(\mathbf{k}_n, \mathbf{k}'_n) \\ &= -2\pi e^2 \Omega \left\{ \frac{c_0^2(\mathbf{k}_n)c_0^2(\mathbf{k}'_n)}{|\mathbf{k}_n - \mathbf{k}'_n|^2} + \frac{c_m^2(\mathbf{k}_n)c_m^2(\mathbf{k}'_n)}{|(\mathbf{k}_n - \mathbf{k}'_n) - 2\pi(\mathbf{m} - \mathbf{m}')/a|^2} \right. \\ &\quad + \frac{c_0^2(\mathbf{k}_n)c_m^2(\mathbf{k}'_n)}{|\mathbf{k}_n - \mathbf{k}'_n + 2\pi\mathbf{m}'/a|^2} + \frac{c_0^2(\mathbf{k}'_n)c_m^2(\mathbf{k}_n)}{|(\mathbf{k}_n - \mathbf{k}'_n) - 2\pi\mathbf{m}/a|^2} \\ &\quad \left. + \frac{2c_0(\mathbf{k}_n)c_0(\mathbf{k}'_n)c_m(\mathbf{k}_n)c_m(\mathbf{k}'_n)}{|\mathbf{k}_n - \mathbf{k}'_n - \pi(\mathbf{m} - \mathbf{m}')/a|^2} \delta \right\}, \quad \dots \quad (4.8) \end{aligned}$$

where $\delta = 1$ if $|\mathbf{m}| = |\mathbf{m}'|$ and 0 otherwise. The total interchange interaction energy is given by

$$J = \sum_n \sum_{n'} J_{nn'} = \left(\frac{\Omega}{8\pi^3} \right)^2 \iint J(\mathbf{k}_n, \mathbf{k}'_n) d\tau_n d\tau'_{n'}, \quad \dots \quad (4.9)$$

the integrations being over the occupied volume in momentum space for each direction of spin (*cf.* (2.10)). The difficulties involved in carrying out such integrations are apparent, having regard to the complexity of (4.8), especially if \mathbf{k}_n or \mathbf{k}'_n is near a zone boundary, and also to the possibly very complicated shape of the Fermi surface.

(b) Approximation of tight binding (Bloch 1928, *cf.* also Sommerfeld and Bethe, *loc. cit.*). This method, which has frequently been regarded as giving the closest approximation in many cases, probably provides the most promising and straightforward approach to the problem of calculating interchange interaction energies for the holes in the d band of transition metals.

The wave functions are here of the form

$$\psi_n(\mathbf{r}_i) = N^{-1/2} \sum_{\mathbf{l}} \exp(i\mathbf{k}_n \cdot \mathbf{R}_l) \chi(\mathbf{r}_i - \mathbf{R}_l), \quad \dots \quad (4.10)$$

where the χ are normalized atomic wave functions and the summation is over all lattice points \mathbf{R}_l . If the electrons are other than of the s type it is necessary to consider the various degenerate m states, to each of which there corresponds a particular atomic wave function χ . Calculations of the density of states in the d band of body-centred cubic metals have been carried out on these lines by Jones and Mott (1937), an extension of whose work to the ferromagnetic metals may eventually do much to clarify the position if, in addition, calculations of interchange interaction energies could be carried out for these metals. All that can be done here is to indicate the kind of relation obtainable with the tight binding approximation; for simplicity interaction between s electrons will be considered, for which the wave functions have the form (4.10) without

further modification (*cf.* also Brillouin 1933). In the present approximation the interatomic distance is assumed large compared with the spread of the atomic wave functions, χ , whose overlap may be neglected for other than neighbouring atoms. On evaluating (2.9) with wave functions (4.10) it is found that, apart from terms which are of the second and higher orders of small quantities, the interchange interaction energy for two electrons with parallel spins and momenta \mathbf{k}_n and \mathbf{k}'_n is given by

$$J_{nn'} = J(\mathbf{k}_n, \mathbf{k}'_n) = -\Sigma [B(\boldsymbol{\rho}_l) + C(\boldsymbol{\rho}_l) \exp \{i(\mathbf{k}_n - \mathbf{k}'_n) \cdot \boldsymbol{\rho}_l\}]. \quad (4.11)$$

Here $\boldsymbol{\rho}_l$ is the distance between neighbouring atoms, $B(\boldsymbol{\rho}_l)$ and $C(\boldsymbol{\rho}_l)$ have the form of molecular Coulomb and exchange integrals, with $C(\boldsymbol{\rho}_l)$ in particular, given by

$$C(\boldsymbol{\rho}_l) = \frac{e^2}{2} \iint \frac{|\chi(\mathbf{r}_i)|^2 |\chi(\mathbf{r}'_i - \boldsymbol{\rho}_l)|^2}{|\mathbf{r}_i - \mathbf{r}'_i|} d\tau_i d\tau'_i,$$

and the summation is over the positions of neighbouring atoms. The relation (4.11) is formally very similar to that derivable for the kinetic energy of tightly bound electrons (*cf.* Sommerfeld and Bethe, *loc. cit.*) and, owing to the occurrence in these expressions of molecular integrals, the conditions for ferromagnetism may finally be formulated in terms of the relative magnitudes of such integrals, so that there is some similarity with the conclusions derivable from the Heitler-London-Heisenberg approximation (*cf.* also § 6).

For the calculation of the total interchange interaction energy, $J(\mathbf{k}_n, \mathbf{k}'_n)$ has again to be integrated over the occupied region of momentum space (*cf.* (4.9) and (2.10)), and, as in the case of nearly free electrons, an accurate result can only be obtained if the distribution of \mathbf{k} states and the shape of the Fermi surface are known with reasonable precision. It appears from this that a necessary preliminary to the calculation of interchange interaction energies of particular metals is the calculation of their electronic energy bands.

(c) Cellular approximation (Wigner and Seitz 1933, 1934, Slater 1934 b, *cf.* also Seitz 1940). This method, which is by far the most complicated, has been applied to the calculation of energy bands of several metals, including nickel (Slater 1936) and iron (Manning 1943, Greene and Manning 1943). The calculation of interchange interaction energies in this approximation presents such difficulties, however, that only very preliminary consideration has been given to it, by Slater (1936), whose calculations suggest that the ferromagnetism of nickel is due almost entirely to the exchange energy associated with the electrons in each particular cell, *i. e.* with the separate nickel atoms. It is difficult to see whether this conclusion, which is quite incompatible with the co-operative nature of ferromagnetism, is the result of inherent shortcomings of the cellular approximation, or whether it could be modified by carrying out more detailed calculations of interchange interaction energies by this method,

The discussion of this section has indicated that to calculate interchange interaction energies of particular metals it is necessary to enter very deeply into details of their electronic constitution, but that it may eventually be possible to obtain reliable results by suitable approximations.

§ 5. DISCUSSION OF SOME EXPERIMENTAL RESULTS.

In the previous paper (Wohlfarth 1949 a) a detailed discussion is given, on the basis of Stoner's treatment, of the magnetic properties of nickel and its alloys above the Curie point. In ordinary fields the spontaneous magnetization is here small ($\zeta \ll 1$), and the total interchange interaction energy very nearly proportional to ζ^2 ; in other words, if the magnetization dependent part of J is given by $-\frac{1}{2}Nk\theta'\zeta^2$, the coefficient θ' should be very nearly independent of temperature. (It will vary with temperature for a number of other reasons, of a secondary character, as was discussed previously.) The experimental results are in good agreement with curves obtained theoretically and enable values of θ' to be estimated unambiguously. Using (3.7), it is, in addition, possible to estimate the relative magnetization at absolute zero, and values derived in this way for nickel and nickel-copper alloys are in good general agreement with values estimated from the observed Curie temperatures.

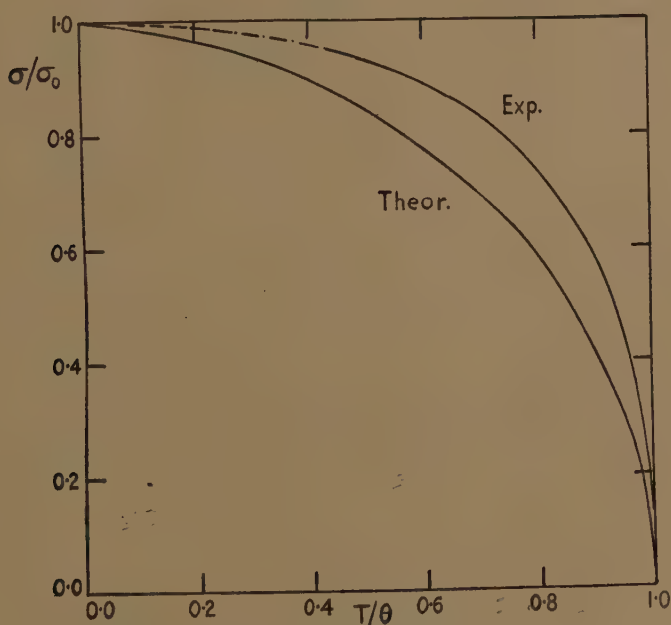
The agreement between theory and experiment is, however, very much less satisfactory below the Curie point. In fig. 2 the upper curve gives the experimentally observed variation of the spontaneous magnetization (written in the usual reduced form σ/σ_0 , where σ_0 is the saturation magnetization) with temperature (written as T/θ , where θ is the Curie temperature) for nickel. The results are due to Weiss and Forrer (1926) and have been extrapolated to zero field by Stoner (1936). The lower curve in fig. 2 is that calculated by Stoner (1938), with $k\theta'/\epsilon_0 = 2^{-1/3}$, the value deduced from the susceptibility results (cf. (3.7)).

The discrepancy may be interpreted on the basis of the discussion of § 3. On increasing the temperature from absolute zero to the Curie point, ζ varies between ζ_0 (for nickel very nearly equal to 1) and 0. Hence the form of the σ - T curve is determined by the variation of the interchange interaction energy over the *whole* magnetization range, while that of the susceptibility, temperature curve depends only on the variation for ζ small. As shown in § 3 (cf. fig. 1), while in the latter case the assumption of a variation of the total $J(\zeta)$ as ζ^2 is justified, the combined effect of parallel and anti-parallel spin interactions will, in general, lead to the occurrence of higher terms in the energy expression than the second when $\zeta^2 \rightarrow 1$. Put differently, the interchange interaction coefficient, θ' , will, in general, vary with temperature below the Curie point. These considerations, although based on the behaviour of free electrons, will apply in a qualitative way to electrons in actual metals and alloys also, and fig. 2 shows that, for nickel, θ' seems to decrease with increasing temperature, giving an increase of the slope of the total interchange interaction curve with ζ^2 (cf. fig. 1). This deviation from strict linearity

will have a marked effect on other properties of ferromagnetics below the Curie point, such as specific heat and the magneto-caloric effect (*cf.* Stoner 1936).

The discussion shows how a careful analysis of experimental results for particular ferromagnetics may enable estimates to be made of the values of interchange interaction energies and of their dependence on the spontaneous magnetization. It is this kind of information which is derivable by theoretical calculations of the type considered in § 4, and when these are finally carried to a satisfactory stage it should be possible to attain good general agreement between theory and experiment for a wide variety of ferromagnetic properties.

Fig. 2.



Magnetization-temperature curve for nickel.

Upper curve : Observed (Weiss and Forrer, 1926).

Lower curve : Calculated from Stoner's treatment.

§ 6. CONCLUDING REMARKS.

It is appropriate, in concluding this paper, to comment briefly on the relative merits of the collective electron treatment of ferromagnetism, and that based on the Heitler-London-Heisenberg approximation. A fairly full discussion of both these treatments has recently been given by Van Vleck (1945), who is of the opinion that it is not possible at present to decide which of these two approaches provides the better approximation. This conclusion, is, however, based partly on a misunderstanding of the assumptions underlying Stoner's treatment. Van Vleck regards the

interchange interaction term in this treatment merely as a phenomenological representation of the Weiss molecular field energy, and believes the "correlation energy" to have been neglected. As shown in the present paper, however, it is just these correlation terms which give rise to the "molecular field".

The Heitler-London-Heisenberg treatment, while originally of great value in interpreting the origin of the molecular field, cannot now be regarded as providing the best approach generally to the quantitative interpretation of the physical properties of ferromagnetic metals; this for a number of reasons. Firstly, a necessary condition for ferromagnetism in this treatment is the occurrence of a positive molecular exchange integral for d wave functions (setting apart the question of the ferromagnetism of gadolinium). This exchange integral has now been calculated (Wohlfarth 1949 b), using spherically symmetrical d functions, and it is found to be negative. Hence the well-known simple qualitative discussion of the conditions for ferromagnetism (see, for example, Sommerfeld and Bethe 1933), based essentially on a consideration of the overlap of radial d wave functions, must be regarded as incorrect. Secondly, the Bloch "spin wave" treatment, frequently regarded as the most rigorous and useful extension of Heisenberg's treatment, has been criticized (Wohlfarth 1949 c) as being based on premises which are probably too idealized for the results to be applicable to actual ferromagnetics, and this conclusion is borne out by a consideration of experimental results. Thirdly, it has, in fact, been recently suggested (Mott 1949) that the Heitler-London method is an unjustifiable approximation in metals generally, and that the collective electron approach should be regarded as providing the best approximation in this case. It is this approach which has been considered in this paper, and it has been shown how it may, in principle, be used to interpret the properties of particular ferromagnetics.

I am indebted to Professor E. C. Stoner, F.R.S. for much helpful discussion; to Mr. P. Rhodes for assistance in the preparation of this paper; and to the Department of Scientific and Industrial Research for a grant.

SUMMARY.

A preliminary discussion is given of the interchange interaction energy terms occurring in the collective electron treatment of ferromagnetism. The theoretical background is outlined (§ 2) and the relevant energy terms calculated for free electrons (§ 3). It is shown that, if the interchange interaction energy for anti-parallel spins is taken into account qualitatively, the total energy may vary as the square of the magnetization, as is assumed in Stoner's statistical treatment (1938). Application to electrons in actual metals will require some very complicated calculations; a brief outline is given (§ 4) of the three main approximational methods which may be used. Valuable information may, however, be obtained by careful

analysis of experimental results (§ 5); a possible explanation may, for example, be given of the discrepancy between the calculated and observed magnetization, temperature curve of nickel.

It is finally suggested (§ 6) that the collective electron treatment provides a much more satisfactory approach to the problems of ferromagnetism than the Heitler-London-Heisenberg method.

REFERENCES.

- BLOCH, F., 1928, *Z. Phys.*, **52**, 555; 1929, *Ibid.*, **57**, 545.
 BRILLOUIN, L., 1932, *J. de Phys.*, **3**, 380, 565; 1933, *Ibid.*, **4**, 1, 333; 1934, *Les champs self-consistents de Hartree et de Fock* (Paris: Hermann).
 GREENE, J. B., and MANNING, M. F., 1943, *Phys. Rev.*, **63**, 203.
 JONES, H., and MOTT, N. F., 1937, *Proc. Roy. Soc. A.*, **162**, 49.
 MANNING, M. F., 1943, *Phys. Rev.*, **63**, 190.
 MOTT, N. F., 1935, *Proc. Phys. Soc.*, **47**, 571; 1949, *Ibid.*, (in the press).
 MOTT, N. F., and JONES, H., 1936, *Theory of the properties of metals and alloys* (Oxford: Clarendon Press).
 PEIERLS, R., 1930, *Ann. Phys. Lpz.*, **4**, 121.
 SEITZ, F., 1940, *Modern Theory of Solids* (New York: McGraw Hill).
 SLATER, J. C., 1934 a, *Rev. Mod. Phys.*, **6**, 209; 1934 b, *Phys. Rev.*, **45**, 794; 1936, *Ibid.*, **49**, 537.
 SOMMERFELD, A., and BETHE, H., 1933, *Handbuch der Physik* (2nd ed.), **24/2**, 333.
 STONER, E. C., 1930, *Proc. Leeds Phil. Soc.*, **2**, 50; 1936, *Phil. Trans. Roy. Soc. A.*, **235**, 165; 1938, *Proc. Roy. Soc. A.*, **165**, 372; 1939, *Ibid.*, **169**, 339; 1948, *Phys. Soc. Rep. Progr. Phys.*, **11**, 43.
 VAN VLECK, J. H., 1945, *Rev. Mod. Phys.*, **17**, 27.
 WEISS, P., and FORRER, R., 1926, *Ann. Phys., Paris*, **5**, 153.
 WIGNER, E., 1934, *Phys. Rev.*, **46**, 1002.
 WIGNER, E., and SEITZ, F., 1933, *Phys. Rev.*, **43**, 804; 1934, *Ibid.*, **46**, 509.
 WOHLFARTH, E. P., 1948, *Proc. Leeds Phil. Soc.*, **5**, 89; 1949 a, *Proc. Roy. Soc. A.*, **195**, 434; 1949 b, *Nature, Lond.*, **163**, 57; 1949 c, *Proc. Leeds Phil. Soc.* (in the press).

LXIX. On the Interaction between Mesons.

By H. Y. TZU,
University of Manchester*.

[Received April 4, 1949.]

ABSTRACT.

The interaction between mesons transmitted by the sea of nucleons in negative energy states is investigated. For low energy mesons, the interaction appears as terms added to the Lagrangian for the meson field *in vacuo*. Weisskopf's (1936) method fails to give the coefficient of the term depending on the fourth power of the field strength, which is mainly responsible for the scattering of mesons by mesons. Instead, the subtraction method of Dirac (1934) and Heisenberg (1934) is used. The

* Communicated by L. Rosenfeld.

range of the interaction is found to be the Compton wavelength of the meson and the singularity at small distances is of the type R^{-5} . The total cross-section for the scattering of mesons by mesons is of the order 10^{-29} cm.² and the existence of a closed stationary or quasi-stationary assembly of mesons is unlikely.

INTRODUCTION.

ACCORDING to Dirac, the vacuum contains an infinite number of protons and neutrons in negative energy states. Two mesons can be absorbed by these nucleons and then re-emitted in different directions; in this way mesons can be scattered by each other. The scattering of meson by meson suggests the existence of an interaction between them, which might eventually lead to the binding together of several mesons in some quasi-stationary state and gives an explanation to the existence of the heavier mesons discovered recently. It is therefore interesting to investigate the properties of the interaction between mesons which is transmitted by nucleons in the negative energy states.

The aim of this paper is to study the general features of this interaction. It begins with a general discussion about the range of the force and the singularity at small distances. A general relation between the form of the singularity of the interaction and the momentum dependence of the scattering matrix element is obtained. Dirac's hole theory is used to derive modified equations for the meson field *in vacuo* valid for processes involving only low energy mesons, which are then applied to the calculation of the cross-section for the scattering of slow mesons by each other. The result shows that the range of interaction is of the same order of magnitude as the Compton wavelength of the meson. The cross-section is of the order of 10^{-29} cm.². Judging from the smallness of the cross-section, the existence of stationary or quasi-stationary configurations of a system of mesons is unlikely. For the scattering of high energy mesons only the momentum dependence of the scattering matrix element is investigated. It is found that the scattering matrix element increases as the square of the momentum of the scattering particles, and it is concluded that the singularity at the origin is of the order R^{-5} , where R is the distance between two mesons.

For the sake of simplicity, the meson field used in the following investigation is neutral and pseudoscalar. Most of the general information obtained about the nature of the interaction between two neutral pseudoscalar mesons can also be applied to other kinds of mesons.

GENERAL CONSIDERATION ABOUT THE RANGE AND THE SINGULARITY OF THE INTERACTION.

The interaction between two mesons is transmitted through the presence of nucleons in negative energy states. At first sight it would be expected that the range of the interaction should be of the order of the Compton wavelength of the nucleon, which is 2×10^{-14} cm. Since this length is

Expression (3) can be reduced symbolically to the form

$$V_n(x) \sim \frac{1}{R^{n+3}} \int \exp\left(-\frac{i}{\hbar c} pR\right) (pR)^n d(\mathbf{p}R), \quad . . . \quad (4)$$

where R is the distance between two interacting particles. The integral is dimensionless apart from a power of $(\hbar c)$. (4) expresses the fact that the stronger the momentum dependence of the scattering matrix element, the stronger is the singularity of the interaction. Denoting the order of the singularity by m , we have then the general relation

$$m-n=3 \quad (5)$$

Together with singularities of the ordinary type R^{-m} , singularities of the δ -function type often occur. These, however, are usually discarded as having no physical significance. The matrix element for Coulomb scattering is, for example, proportional to

$$\sim p^{-2} \quad (6)$$

From (5) it is immediately deduced that the singularity is of the type R^{-1} . Similar argument can be used to prove that the potential between two nucleons in the conventional meson theory is of the type R^{-3} . The potential actually calculated also includes terms of the type

$$\sim \frac{1}{R} \delta'(R) \quad (7)$$

which are usually discarded.

Heisenberg (1936) has pointed out that, in a theory involving a constant of the dimensions of a length, processes of an explosive type may be expected to occur in the high energy region. In such a theory the cross-section increases with the momentum in many processes. While such processes are rare in electromagnetic phenomena, they are very common in nuclear phenomena. It is to be expected that the singularity of the interaction at small distances between two mesons will turn out to be strong.

THE SCATTERING MATRIX ELEMENT.

This section is a rough account of those parts of the meson field theory of nuclear forces, which will be used in the following discussions. The Lagrangian

$$L=L_n+L_m+L_i \quad (8)$$

used in such a theory is made of three parts. The first part

$$L_n=-\Psi^*(x^\nu)\{\alpha^\mu p_\mu+\beta M\}\Psi(x^\nu) \quad (9)$$

which is constructed entirely out of variables describing the nucleons, is due to the free nucleon field. Ψ is the quantized nucleon wave function, α^μ , β are the Dirac matrices; and M the mass of the nucleon. x^ν denote the space and the time variables with the suffix ν denoting 0, 1, 2, 3. The second part

$$L_m=-\frac{1}{2}\{G_\mu G^\mu+\kappa^2 U^2\}, \quad (10)$$

According to perturbation theory, the matrix element for the mutual scattering of two mesons is of the form

$$H_{sc} = \sum \frac{H_{b\text{III}} H_{\text{IIIII}} H_{\text{IIII}} H_{Ia}}{(E_a - E_I)(E_a - E_{\text{II}})(E_a - E_{\text{III}})}, \quad \dots \quad (15)$$

where a and b denote the initial and the final states respectively. I, II, III denote the three intermediate states required for the absorption and the re-emission of two mesons. H_{Ia} , H_{IIII} , H_{IIIII} and $H_{b\text{III}}$, which represent the four steps of the transition, are to be derived from (14) in the usual way. The summation is to include :

- (a) Processes in which one, two or three nucleons take part.
- (b) The different orders in which the absorption and the emission of the mesons take place.
- (c) Spin states of the nucleon.
- (d) The different momentum states of a nucleon in the negative energy states.
- (e) The proton and the neutron states.

To obtain quantitative information directly from (15) is very difficult. Firstly, the calculation of such a matrix element is very tedious. Also, as can be seen easily, the summation over the infinite number of the momentum states of the nucleon causes the matrix element to diverge logarithmically, always provided the terms of highest power in the momentum \mathbf{p} of the nucleon do not cancel each other. Therefore some way of cutting off the infinity and of simplifying the calculation must be found. In the next section, the subtraction methods developed by Dirac, Heisenberg and Weisskopf will be used to deal with low energy scattering. For high energy scattering only qualitative information is aimed at. This is discussed in the later part of the paper.

THE MODIFIED HAMILTONIAN OF THE MESON FIELD IN VACUO AND THE SCATTERING OF LOW ENERGY MESONS.

The Hamiltonian derived from (8) is

$$H = \frac{1}{2} \left\{ \mathbf{G}^2 + G_0^2 + \kappa^2 U^2 \right\} + \Psi^* \left\{ \alpha^i p_i + \beta M + \frac{f}{\kappa} \boldsymbol{\sigma} \cdot \mathbf{G} \right\} \Psi. \quad (16)$$

The suffix i takes the values 1, 2, 3, only. In order to take into account Dirac's hole theory, the Lagrangian L_m of the meson field in the vacuum should be modified in such a way that the Hamiltonian derived from it will include, besides the energy density of the free meson field, which is the first part of the expression (16), also the contribution from nucleons in the negative energy states represented by the second part of (16).

In Hartree's approximation, the equation of motion of a single nucleon derived from (9) and (12) is

$$\left\{ \alpha^\mu p_\mu + \beta M + \frac{f}{\kappa} \boldsymbol{\sigma} \cdot \mathbf{G} + \frac{f}{\kappa} \rho_1 G_0 \right\} \psi = 0. \quad (17)$$

Let the exact Lagrangian for the meson field in vacuum be written as

$$L_v = L_m + L_s \quad (18)$$

where L_m is the zero order approximation (10) and L_s the modification. Having put g equal to zero, (17) and the second part of (16) depend no longer on the field potential U . It should be expected that L_s will also not depend on U . On the other hand, owing to the low energy of the mesons, the variation of the field strength G_μ in space and time is small, so that L_s can be regarded as being independent of the derivatives of the field strength. To this approximation, L_s is a function of G_μ alone. The additional Hamiltonian should then be

$$H_s = G_0 \frac{\partial L_s}{\partial G_0} - L_s = \sum_n \psi_n^* \left\{ \alpha^i p_i + \beta M + \frac{f}{\kappa} \boldsymbol{\sigma} \cdot \mathbf{G} \right\} \psi_n, \quad (19)$$

where n is to be summed over all the negative energy states. Owing to the minimum character of the energy of the state ψ_n , the right-hand side of (19) is equal to

$$G_0 \frac{\partial}{\partial G_0} \left[- \sum_n \psi_n^* \left\{ \alpha^i p_i + \beta M + \frac{f}{\kappa} \boldsymbol{\sigma} \cdot \mathbf{G} + \frac{f}{\kappa} \rho_1 G_0 \right\} \psi_n \right] \\ - \left[- \sum_n \psi_n^* \left\{ \alpha^i p_i + \beta M + \frac{f}{\kappa} \boldsymbol{\sigma} \cdot \mathbf{G} + \frac{f}{\kappa} \rho_1 G_0 \right\} \psi_n \right]. \quad (20)$$

From (17), (19) and (20) it follows

$$L_s = -i\hbar c \sum_n \psi_n^* \frac{\partial}{\partial x_0} \psi_n \quad (21)$$

It is easy to see from equation (17) that the quantities available for use in constructing L_s are the following:

$$B_\mu = \frac{f}{\kappa} G_\mu, \quad \hbar, c, M. \quad (22)$$

If L_s is an analytic function of B_μ and can therefore be developed in power series for weak B_μ , then from the dimensional consideration and the consideration of Lorentz invariance it is easy to see that L_s must be of the form

$$L_s = - \sum_n A_n \frac{M^4}{\hbar^3 c^3} \left(\frac{B_\mu B^\mu}{M^2} \right)^n, \quad (23)$$

where A_n are dimensionless numerical coefficients. The term which is

mainly responsible for the scattering of mesons by mesons, is

$$-\frac{A_2(B_\mu B^\mu)^2}{\hbar^3 c^3} = -\frac{A_2 f^4 (G_\mu G^\mu)^2}{\hbar^3 c^3 \kappa^4} \quad (24)$$

It is interesting to note that the only quantity describing the nucleon which appears in (24) is the interaction constant f . The mass of the nucleon does not enter this expression. Thus when treating the scattering of mesons by mesons with (24) the Compton wavelength of the nucleon does not appear in the expression for the cross-section. This indicates that there is no relation between the range of the interaction between mesons and the mass of the nucleon. This result arrived at by the dimensional argument is quite general and is also valid for other kinds of meson fields, so long as :

(1) The mass of the meson is small and therefore its Compton wavelength is large in comparison with those of the nucleon.

(2) The energy of the mesons involved in the process is small in comparison with the rest mass of the nucleon.

It is also interesting to compare the result with that obtained by Euler and Kockel (1935) for the electromagnetic field. Putting $\hbar c \kappa = m$, the mass of the meson, (24) becomes

$$-A_2 \frac{f^4 \hbar c}{m^4} (G_\mu G^\mu)^2, \quad (25)$$

while the corresponding terms for the electromagnetic field are

$$\frac{e^4 \hbar c}{360 \pi^2 m_e^4} [(\mathbf{E}^2 - \mathbf{H}^2)^2 + 7(\mathbf{E} \cdot \mathbf{H})^2], \quad (26)$$

where e , m_e , \mathbf{E} and \mathbf{H} are the electronic charge, the electronic mass, the electric field strength and the magnetic field strength respectively. (25) would be equivalent to (26) if the mass of the meson were replaced by M , the mass of the nucleon. It will be found that the analogy between the result obtained for meson and that for photon is very limited. The difference is due to the different nature of these two fields and the different form of the interaction terms. For the meson field a constant field potential has a physical meaning, a constant field potential in the electromagnetic field has no meaning. While there are interaction terms between the meson field and the nucleons which depend on the meson field strength, there is no such interaction term between the electromagnetic field and the electrons.

According to Weisskopf (1936), the coefficients A_n may be determined by solving (17) for a particular field. The wave function thus obtained will be put into (21). L_s will then come out as a function of this particular field. By developing this function into a power series in the field strength, it will be found that both the field independent term and the term depending on the second power of the field strength diverge. Since a field independent term has no bearing on the equation of motion and a

The modified Hamiltonian for the meson field *in vacuo* then becomes :

$$H_v = \frac{1}{2} \{ \mathbf{G}^2 + \tilde{\mathbf{G}}_0^2 + \kappa^2 \mathbf{U}^2 \} - \frac{f^4}{6\pi^2 \hbar^3 c^3 \kappa^4} \{ \mathbf{G}^2 - \tilde{\mathbf{G}}_0^2 \}^2 + \dots \quad (34)$$

with
$$\tilde{\mathbf{G}}_0 = \mathbf{G}_0 - \frac{2f^4 \mathbf{G}_0 (\mathbf{G}_\mu \mathbf{G}^\mu)}{3\pi^2 \hbar^3 c^3 \kappa^4} \quad (35)$$

The scattering matrix element of two colliding mesons can be calculated in the usual way from (34). After the quantization, the first part of H_v gives the energy of free mesons, while the second part gives rise to the scattering process. In a coordinate system in which the centre of gravity of the two mesons is at rest, the total cross-section is

$$\phi = \frac{f^8}{72\pi^5 (\hbar c)^{10} \kappa^8 E^2} \left\{ 9E^8 + 12E^6 p^2 + 14E^4 p^4 + \frac{20}{3} E^2 p^6 + \frac{47}{15} p^8 \right\}, \quad (36)$$

where p and E are the momentum and energy of the individual meson respectively. For slow mesons, the terms depending on p can be neglected and the energy E can be equated to the rest energy of the meson. (36) becomes then

$$\phi = \frac{32}{\pi} \left(\frac{f^2}{4\pi \hbar c} \right)^4 \left(\frac{1}{\kappa} \right)^2 \quad (37)$$

The structure of the expression (37) for the total cross-section for the scattering of slow mesons is simple. It is the square of the Compton wavelength of the meson multiplied by a dimensionless factor. The interpretation of (37) is straightforward. The dimensionless factor represents the effect of an interaction constant between two mesons. The factor $(1/\kappa)^2$ represents the effect of the range of the force, which should be therefore of the order of the Compton wavelength of the meson.

Putting

$$\frac{f^2}{4\pi \hbar c} = 0.07, \quad \frac{1}{\kappa} = 2 \times 10^{-13} \text{ cm.}, \quad (38)$$

the total cross-section ϕ turns out to be only 10^{-29} cm.^2 , which is very much smaller than that for the scattering of nucleons. Since the range of the interaction between mesons is of the same order as that between nucleons, it suggests that the strength of interaction between mesons is far weaker than that between nucleons.

The possibility of the existence of any stationary or quasi-stationary state can be estimated from the following consideration. Assuming that the average distance between two mesons forming a system is of the order of the range of the interaction $1/\kappa$, the kinetic energy of the system is of the order $\hbar c \kappa$. On the other hand, the potential energy of the system is of the order of magnitude

$$\hbar c \kappa \sqrt{\left\{ \frac{32}{\pi} \left(\frac{f^2}{4\pi \hbar c} \right)^4 \right\}} = 0.015 \hbar c \kappa,$$

which is only about 1 per cent of the kinetic energy. The existence of any stationary or quasi-stationary state is then very unlikely. However,

if the singularity of the interaction is of the attractive type and sufficiently strong, then, as the average distance between the mesons is decreased the potential energy increases far more rapidly than the kinetic energy, and a state of binding might eventually be formed. But, in order to have a finite binding energy it is necessary to cut-off the singularity at small distances.

It is interesting to compare the above result with that obtained by Euler (1936) when he treated the corresponding problem for the electromagnetic field. The total cross-section for the scattering of light by light in the centre of gravity system is

$$\phi = \frac{224\pi}{10125} \left(\frac{\lambda_0}{\lambda} \right)^8 \lambda^2, \quad \lambda_0 = \frac{\hbar c}{m_e} \sqrt{\left(\frac{e^2}{\hbar c} \right)}, \quad \dots \quad (40)$$

where λ is the wavelength of the light. The form of (40) is quite different from that of (37). There is no question of the range of the interaction, since no meaning can be attached to the position of a photon. At the same time the concept of a static interaction must be abandoned. Thus the cross-section (40) vanishes as the momentum of the photon tends to zero.

THE SCATTERING OF THE HIGH ENERGY MESONS.

A quantitative treatment of the scattering of high energy mesons is very complicated. The following discussion is therefore only qualitative. The purpose of this section is to find out the momentum dependence of the scattering matrix element and thereby determine the type of singularity of the interaction at small distances between two mesons. It is necessary to start directly from (15).

By transforming the summation over the momenta of the nucleons into an integral, (15) becomes

$$H_{sc} = \frac{V}{(2\pi\hbar c)^3} \int_0^\infty d\mathbf{p} \sum \frac{H_{b\text{III}} H_{\text{III}\text{II}} H_{\text{II}\text{I}} H_{\text{I}a}}{(E_a - E_{\text{I}})(E_a - E_{\text{II}})(E_a - E_{\text{III}})} \dots \quad (41)$$

The integral diverges logarithmically as the upper limit of \mathbf{p} tends to infinity. This prevents quantitative information being drawn immediately from (41). However, qualitative information can still be obtained without the help of any detailed method of cutting off the infinity.

It can be safely assumed that the main contribution to the scattering of mesons of momenta of the order of p_0 must come from those nucleons in negative energy states, whose momentum vectors lie within a sphere in the momentum space whose radius is comparable with p_0 . Let the radius be μp_0 where μ is a number of the order of 1, presumably larger than 1. Then as a rough but qualitatively correct approximation, the upper limit of the integral (41) can be replaced by μp_0 .

Consider a scattering process with a certain momentum diagram for the mesons. Let the momenta have the order of magnitude of p'_0 . The scattering matrix element is then

$$H'_{sc} \simeq \frac{V}{(2\pi\hbar c)^3} \int_0^{\mu p'_0} d\mathbf{p}' \sum \frac{H_{b'\text{III}'} H_{\text{III}'\text{II}'} H_{\text{II}'\text{I}'} H_{\text{I}'a'}}{(E_{a'} - E_{\text{I}'})(E_{a'} - E_{\text{II}'})(E_{a'} - E_{\text{III}'})} \dots \quad (42)$$

Imagine another scattering process, whose momentum diagram is exactly similar to that of the first process, but magnified ν times. The corresponding scattering matrix element is

$$H''_{sc} \cong \frac{V}{(2\pi\hbar c)^3} \int_0^{\mu p''} d\mathbf{p}'' \Sigma \frac{H_{b''}{}_{III''} H_{III''}{}_{II''} H_{II''}{}_{I''} H_{I''}{}_{a''}}{(E_{a''} - E_{I''})(E_{a''} - E_{II''})(E_{a''} - E_{III''})}. \quad (43)$$

To find out the relation between (42) and (43), the substitution

$$p'' = \nu p', \quad \dots \dots \dots (44)$$

will be made. In a high energy scattering process, the rest mass of the meson and that of the nucleon can be neglected in comparison with p' and p'' . To this approximation,

$$H_{I''}{}_{a''} \cong \sqrt{(\nu)} H_{I'}{}_{a'}, \quad E_{a''} \cong \nu E_{a'}, \dots \text{etc.} \quad (45)$$

From (42), (43), (44) and (45) it follows that

$$H''_{sc} \cong \nu^2 H'_{sc}. \quad \dots \dots \dots (46)$$

The scattering matrix element increases therefore as the square of the momentum, always provided that the terms of the highest power in momentum do not cancel each other exactly. From (5) it follows that the interaction between two mesons has a singularity of the type

$$\sim \frac{1}{R^5}, \quad \dots \dots \dots (47)$$

at small distances.

It is interesting to compare this result with that of the scattering of high energy photons treated by Achieser (1937). He obtained a matrix element decreasing as the inverse of the square of the momentum. The cause of this difference between the momentum dependence of the scattering of mesons and that of the scattering of photons is evident. While the expression (14) for the absorption and the emission of a meson increases as the square root of the momentum, the corresponding expression for the absorption and the emission of a photon decreases as the inverse of the square root of the momentum. This is in turn due to the fact that while the electron field is only coupled with the electromagnetic field potential, the nucleon field is coupled with the derivatives of the meson potential as well as with the potential itself. Equation (47) is therefore a general result; its validity is not restricted to the pseudoscalar field, as it can also be applied to other kinds of meson fields, so long as both the field potential and its derivatives are introduced into the interaction terms.

The above treatment has not allowed for the resonance phenomena, which would appear as soon as the energy of the mesons is sufficiently high to produce nucleon pairs. In particular, the factor

$$(E_a - E_{II}) \quad \dots \dots \dots (48)$$

in the denominator of the expression (41) will tend to zero for certain values of the momenta \mathbf{p}_r of the nucleons which occur in the intermediate

CONCLUSION.

From the above discussion it is clear that this interaction between mesons is unlikely to lead to stationary or quasi-stationary states for a cluster of mesons. Even if, owing to some extraordinary form of the interaction, such stationary or quasi-stationary states do exist it would be very difficult to describe them in a quantitative manner. The difficulty is not entirely due to the involved nature of the calculations which are required in finding the interaction energy. Since the range of the interaction is of the same order of magnitude as the Compton wavelength of the meson, the kinetic energy of the mesons in such a closed stationary state must be comparable to their rest masses; and the non-relativistic Schrödinger wave equation is no longer applicable. The problem is therefore of the relativistic many body type which can hardly be solved at present.

From the theoretical point of view, the high order of singularity of the interaction at small distance is of considerable interest. It is doubtful whether such a high order singularity has any physical meaning. Since the singularity is connected with high energy processes, the question of the applicability of the present quantum theory to processes of high energy arises. According to Heisenberg, the present quantum theory is incapable of describing processes taking place at a very small distance, and in such phenomena concepts like "potential energy" etc., have to be abandoned. The present investigation is an indication of what might happen when the present quantum theory is applied to regions beyond the customary limit.

The author would like to express his thanks to Professor Rosenfeld for suggesting the problem and for many valuable discussions.

APPENDIX.

According to Dirac, a system of nucleons can be characterized to Hartree's approximation by the density matrix

$$(x'', s' | R | x'', s'') = \sum_n \psi_n(x'', s') \psi_n^*(x'', s''), \quad . \quad . \quad . \quad (54)$$

where x'', s are the space, time and spin coordinates respectively. From the density matrix all important physical properties of the system can be deduced. For the sake of symmetry, two other matrices are introduced.

$$R_F = \sum_{occ} \psi \psi^* + \sum_{un} \psi \psi^*, \quad R_I = \sum_{occ} \psi \psi^* - \sum_{un} \psi \psi^* \quad . \quad . \quad . \quad (56)$$

where \sum_{occ} is the summation over all occupied states and \sum_{un} the summation over all the unoccupied states. R is then

$$R = \frac{1}{2}(R_F + R_I). \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

All the three matrices satisfy the equations

$$\mathcal{H} R_F = 0, \quad \mathcal{H} R_I = 0, \quad \mathcal{H} R = 0,$$

with

$$\mathcal{H} = -i\hbar c \frac{\partial}{\partial x^0} + H, \quad (58)$$

where H is the Hamiltonian for the individual nucleon.

For the case of no field R_F has been calculated by Dirac as

$$R_{F(\text{vac})} = \frac{i}{\pi\hbar c} [\alpha^\mu p^\mu + \beta M] \frac{1}{2r} \cdot \frac{\partial}{\partial r} U_{F(\text{vac})} \quad (59)$$

with

$$U_{F(\text{vac})} = \sum_{n=0}^{\infty} \frac{\left(\frac{M^2}{4\hbar^2 c^2} x_\mu x^\mu \right)^n}{(n!)^2} \{ \epsilon(x^0 - r) - \epsilon(-x^0 - r) \}$$

$$x^\mu = x^{\mu'} - x^{\mu''}, \quad r^2 = x_i x^i,$$

$$\epsilon(\xi) = 1 \text{ for } \xi > 0,$$

$$\epsilon(\xi) = 0 \text{ for } \xi < 0.$$

If all the states of negative energy are occupied and those of positive energy unoccupied R_1 is given by

$$R_{1(\text{vac})} = \frac{i}{\pi\hbar c} [\alpha^\mu p^\mu + \beta M] \frac{1}{2r} \cdot \frac{\partial}{\partial r} U_{1(\text{vac})}, \quad (60)$$

with

$$U_{1(\text{vac})} = \frac{i}{\pi} \sum_{n=0}^{\infty} \frac{\left(\frac{M^2}{4\hbar^2 c^2} x_\mu x^\mu \right)^n}{(n!)^2} \left\{ \log \frac{\gamma^2 M^2}{4\hbar^2 c^2} |x_\mu x^\mu| - 2\phi(n) \right\}$$

$$\gamma = 1, 7810 \, 72 \dots, \quad \phi(0) = 0.$$

$$\phi(n) = 1 + \frac{1}{2} + \dots + \frac{1}{n} \text{ for } n > 0.$$

While the singularities of $R_{F(\text{vac})}$ are of the forms

$$\delta'(x_\mu x^\mu), \quad \delta(x_\mu x^\mu), \quad \epsilon(x_\mu x^\mu); \quad (61)$$

those of $R_{1(\text{vac})}$ are of the types

$$1/(x_\mu x^\mu)^2, \quad 1/(x_\mu x^\mu), \quad \log |x_\mu x^\mu|. \quad (62)$$

It may be expected that the singularities are still of the same types when a field is present. Assuming a field (27) the corresponding density matrices have been calculated and it is found that

$$\left. \begin{aligned} \Sigma_s(x^{\mu'}, s | R_F(B_0) | x^{\mu''}, s) &= -\frac{4}{\pi} \cdot \frac{\partial}{\partial x^0} \cdot \frac{1}{2r} \cdot \frac{\partial}{\partial r} \left\{ U_{F(\text{vac})} \cos \frac{B_0 r}{\hbar c} \right\} \\ \Sigma_s(x^{\mu'}, s | R_1(B_0) | x^{\mu''}, s) &= -\frac{4}{\pi} \cdot \frac{\partial}{\partial x^0} \cdot \frac{1}{2r} \cdot \frac{\partial}{\partial r} \left\{ U_{1(\text{vac})} \cos \frac{B_0 r}{\hbar c} \right\} \end{aligned} \right\} \quad . . . (63)$$

Their singularities are in fact of the same types as those given in (61) and (62).

The singularities are responsible for the divergence of many results in Dirac's hole theory. Dirac has suggested a procedure to cut off all these singularities. According to Dirac, $R_F(B_\mu)$ is to be removed as a whole from the expression (57) in order to obtain a finite density matrix R . One of the possible expressions which can be cut away from $R_1(B_\mu)$ consistently with Dirac's procedure is

$$R_s(B_\mu) = \exp\left(-i\rho_1 \frac{B_\mu x^\mu}{\hbar c}\right) \cdot R_{1(\text{vac})} + \frac{v}{x_\mu x^\mu} + w \log \frac{\gamma^2 M^2}{4\hbar^2 c^2} |x_\mu x^\mu| \quad . \quad . \quad (64)$$

with

$$\begin{aligned} v = & \frac{iM\beta\rho_1}{\pi^2\hbar^4c^4} \left\{ 2\hbar^2c^2(B_\mu x^\mu) - \hbar^2c^2(\alpha^\mu x_\mu)(\alpha^\nu B^\nu) \right. \\ & - \frac{1}{3}(B_\mu x^\mu)^3 - \frac{1}{2}M^2(B_\mu x^\mu)(x_\nu x^\nu) \\ & + \frac{1}{6}(\alpha^\mu x_\mu)(B_\nu x^\nu)^2(\alpha^e B^e) \\ & \left. + \frac{1}{12}M^2(\alpha^\mu x_\mu)(x_\nu x^\nu)(\alpha^e B^e) + \dots \right\} \\ w = & \frac{iM\beta\rho_1}{\pi^2\hbar^4c^4} \left\{ \frac{i}{2}\hbar c\rho_1(B_\mu B^\mu) - \frac{i}{2}M\hbar c\beta(\alpha^\mu B^\mu) \right. \\ & + \frac{1}{2}M^2(B_\mu x^\mu) + \frac{1}{6}(B_\mu B^\mu)(B_\nu x^\nu) \\ & - \frac{1}{4}M^2(\alpha^\mu x_\mu)(\alpha^\nu B^\nu) - \frac{1}{4}M\beta\rho_1(\alpha^\mu x^\mu)(B_\nu B^\nu) \\ & \left. - \frac{1}{6}(\alpha^\mu x_\mu)(\alpha^\nu B^\nu)(B_e B^e) + \dots \right\}. \end{aligned}$$

From (63) and (64) the modified density matrix can be obtained. The additional Lagrangian L_s can then be derived and A_n determined.

REFERENCES.

- ACHESER, 1937, *Phys. Zeits. d. Sowj.*, **11**, 261.
 DIRAC, 1934, *Camb. Phil. Soc. Proc.*, **30**, 150.
 EULER, 1936, *Ann. d. Phys.*, (5), **26**, 389.
 EULER and KOCKEL, 1935, *Naturwiss.*, **23**, 246.
 HEISENBERG, 1934, *Zeits. f. Phys.*, **90**, 209; 1936, *Ibid.*, **101**, 533; 1939, *Ibid.*, **113**, 61.
 WEISSKOPF, 1936, *Det. Kgl. Dan. Vid. Sel. Mat. Fys.*, **14**, 6.

LXX. *Lattice Theory of Dielectric and Piezoelectric Constants in Crystals.*

By KUN HUANG,

Department of Theoretical Physics, University of Liverpool*.

[Received March 11, 1949.]

ABSTRACT.

A general perturbation theory has been given by Born for treating long acoustic lattice waves. By comparing the wave equations thus obtained with the equations for elastic waves he obtains the expression for the elastic constants. The formal results however diverge when Coulomb interactions between charged lattice particles are taken into account. The fault in this case can be traced to the breakdown of the expansion procedure employed in the perturbation theory. In this paper it is shown that after the separation of a term by Ewald's theta-function transformation from the lattice restoring force, the perturbation theory can once more be developed. The corresponding wave equation contains an extra contribution due to the term separated which is shown to be the macroscopic electric field associated with the wave. The equation proves to be identical with the rigorous elastic wave equation for a piezoelectric crystal which takes into account the interaction between the elastic stresses and the electric field produced by the piezoelectric polarization. The comparison leads to the expressions for the dielectric and piezoelectric coefficients as well as the elastic constants (which cannot be obtained by the standard static methods when Coulomb interaction is taken into account).

§1. INTRODUCTION.

IN the lattice theory, the static elastic, dielectric and piezoelectric effects are usually treated in the following manner (*cf.* Born 1923, Born and Göppert-Mayer 1933). Suppose that an infinite lattice is subject to a given external strain and a homogeneous electric field. If internal strains were zero, *i. e.* if the particles in the basis are merely displaced in accordance with the external strain, forces would in general be experienced by the lattice particles. Therefore suitable additional shifts between particles in the basis (internal strain) are caused to occur which give rise to additional forces on the lattice particles just sufficing to balance the other forces. Once the internal strains are determined in this way, the potential function becomes a function of the external strain and the electric field. The electric polarization and mechanical stresses are similar functions obtained by differentiating the potential function with respect to the electric field and the strain components.

* Communicated by Professor M. Born, F.R.S.

In this theory the electric polarization is ascribed to the displacements of charges carried by lattice particles due to internal strain. However, if actual account is taken of the Coulomb interaction between the charged particles, the internal strain would give rise to a uniform polarization. The macroscopic electric field due to the polarization is completely indeterminate in an infinite lattice and is determined by the boundaries in a finite lattice. Therefore in the infinite lattice model the forces exerted on the charged lattice particles due to the internal strain are accordingly indeterminate. Thus one finds that the formal results in the usual theory diverge when Coulomb interaction is explicitly considered.

In such cases, convergent expressions have been substituted by Born on the basis of various arguments (Born 1923, pp. 728, 773, Born and Göppert-Mayer 1933, pp. 732, 774). The ultimate justification, however, appears to lie in the occurrence of certain terms in the Born-Ewald electromagnetic lattice theory which are similar to the divergent terms in the static theory. The argument is indirect and difficult to present in a close and neat manner. It is the purpose of this paper to show that by the generalization of a perturbation theory due to Born, the expressions for the various coefficients come out in an inevitable way from the mathematical analysis.

The perturbation theory was developed by Born (1923, *cf.* Begbie and Born 1947) for treating long acoustic lattice waves. The equations for very long waves are found to be identical with that for elastic waves. The comparison enables him to identify the expressions for the elastic constants. The results obtained in this way are identical with those obtained by the static method, therefore also divergent when Coulomb interactions are considered. The failure of the method in this case can be traced to the breakdown of the expansion procedure. But turning to the macroscopic point of view, it is noted that in a piezoelectric crystal, polarization will be caused by stresses in an elastic wave so an electric field will be associated with the wave. Therefore one finds that in the rigorous macroscopic equations for the elastic waves electric field explicitly appears. It should thus be expected that if the lattice wave equations are properly set up, terms containing the macroscopic electric field should occur, apart from elastic stresses due to strains. It will be shown that this is actually the case. By the use of Ewald's theta-function transformation, the macroscopic electric field can be separated from the lattice forces. The expansion procedure can then be applied separately to the electric field and the remaining lattice forces in the manner of Born's theory. The long wave equations finally obtained are identical with the rigorous macroscopic equations containing explicitly the electric field. Identifications for the various coefficients follow directly from the comparison. This treatment bears out in a clear way the physical interpretation of the divergence of the static results. The interpretation is already inherent in the justification given by Born in terms of the electromagnetic lattice theory; but as it has not been very forcibly given its significance does not appear to have been generally appreciated,

§ 2. THE WAVE EQUATION.

For quantities relating to the crystal lattice we shall use the following notations :

$l(l^1, l^2, l^3)$	=lattice cell index.
$k(0 \dots n-1)$	=base index (n particles in the basis).
$\alpha, \beta, \gamma \dots$	=indices for Cartesian components.
$\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$	=lattice basic vectors.
$\mathbf{b}^1, \mathbf{b}^2, \mathbf{b}^3$	=basic reciprocal lattice vectors.
m_k	=mass of particle k in the basis.
v_a	=volume of the lattice cell.
$\mathbf{x} \begin{pmatrix} l \\ k \end{pmatrix} = \mathbf{x}(l) + \mathbf{x}(k)$	=lattice point occupied by particle $\begin{pmatrix} l \\ k \end{pmatrix}$.
$\mathbf{x}(l) = l^1 \mathbf{a}_1 + l^2 \mathbf{a}_2 + l^3 \mathbf{a}_3$	
$\mathbf{x} \begin{pmatrix} l-l' \\ kk' \end{pmatrix} = \mathbf{x} \begin{pmatrix} l \\ k \end{pmatrix} - \mathbf{x} \begin{pmatrix} l' \\ k' \end{pmatrix}$	
$\mathbf{u} \begin{pmatrix} l \\ k \end{pmatrix}$	=small displacement vector of $\begin{pmatrix} l \\ k \end{pmatrix}$.
Φ	=lattice energy as function of the displacements.
$\Phi_{\alpha\beta} \begin{pmatrix} l-l' \\ kk' \end{pmatrix} = \left(\frac{\partial^2 \Phi}{\partial u_\alpha \begin{pmatrix} l \\ k \end{pmatrix} \partial u_\beta \begin{pmatrix} l' \\ k' \end{pmatrix}} \right)_0$	=second derivative of the energy function at equilibrium configuration, depending only on the relative cell index.

The equations of motion for small vibrations are

$$m_k \ddot{u}_\alpha \begin{pmatrix} l \\ k \end{pmatrix} = - \sum_{l'k'\beta} \Phi_{\alpha\beta} \begin{pmatrix} l-l' \\ kk' \end{pmatrix} u_\beta \begin{pmatrix} l' \\ k' \end{pmatrix} \dots \dots \dots (2.1)$$

The solution can be written as

$$u_\alpha \begin{pmatrix} l \\ k \end{pmatrix} = \frac{1}{\sqrt{m_k}} w_\alpha \left(k \left| \begin{matrix} y \\ j \end{matrix} \right. \right) \exp \left\{ 2\pi i \mathbf{y} \cdot \mathbf{x} \begin{pmatrix} l \\ k \end{pmatrix} - i \omega \begin{pmatrix} y \\ j \end{pmatrix} t \right\}, \quad \dots \dots (2.2)$$

where \mathbf{y} is an arbitrary wave number vector and $j=0 \dots 3n-1$ specifies the $3n$ independent solutions for given \mathbf{y} . Substitution of (2.2) in (2.1) leads to the following $3n$ linear homogeneous equations :

$$\omega^2 \begin{pmatrix} y \\ j \end{pmatrix} w_\alpha \left(k \left| \begin{matrix} y \\ j \end{matrix} \right. \right) = \sum_{k'\beta} C_{\alpha\beta} \begin{pmatrix} y \\ kk' \end{pmatrix} u_\beta \left(k' \left| \begin{matrix} y \\ j \end{matrix} \right. \right), \quad \dots \dots (2.3)$$

where

$$C_{\alpha\beta} \begin{pmatrix} y \\ kk' \end{pmatrix} = \frac{1}{\sqrt{(m_k m_{k'})}} \sum_l \Phi_{\alpha\beta} \begin{pmatrix} l \\ kk' \end{pmatrix} \exp \left\{ -2\pi i \mathbf{y} \cdot \mathbf{x} \begin{pmatrix} l \\ kk' \end{pmatrix} \right\}. \quad \dots (2.4)$$

There are three solutions of (2.3) with frequencies linearly proportional to the wave number at the long wave limit. These are the *lattice acoustic waves* which become identical with the macroscopic elastic waves at the long wave limit, that is, when $\mathbf{y} \rightarrow 0$. In Born's treatment (Begbie and Born 1947) the solutions for such waves are obtained by expanding the quantities

$$C_{\alpha\beta}\left(\frac{\mathbf{y}}{kk'}\right), \quad w_{\alpha}\left(k\left|\frac{\mathbf{y}}{j}\right.\right), \quad \omega\left(\frac{\mathbf{y}}{j}\right)$$

in powers of \mathbf{y} and the equations of various orders are solved successively following the perturbation procedure. The elastic wave equations are obtained on proceeding to the second order equations. The expansion of $C_{\alpha\beta}\left(\frac{\mathbf{y}}{kk'}\right)$ becomes, however, impossible when the lattice particles are electrically charged. Consider for instance the zeroth order term

$$C_{\alpha\beta}^{(0)}\left(\frac{\mathbf{y}}{kk'}\right) = \frac{1}{\sqrt{(m_k m_{k'})}} \sum_l \Phi_{\alpha\beta}\left(\frac{l}{kk'}\right). \quad \dots \quad (2.5)$$

$-\sum_l \Phi_{\alpha\beta}\left(\frac{l}{kk'}\right)$ could be interpreted as the α -component of the force on $\left(\frac{0}{k}\right)$ when all the k' -particles are displaced by unity in β -direction. When k' has a net charge, the result is completely indeterminate in an infinite lattice. It will be shown in the next section that one can write the restoring force on the right of (2.3) as

$$-\frac{e_k E_{\alpha}}{\sqrt{m_k}} + \sum_{k'} \bar{C}_{\alpha\beta}\left(\frac{\mathbf{y}}{kk'}\right) w_{\beta}\left(k'\left|\frac{\mathbf{y}}{j}\right.\right), \quad \dots \quad (2.6)$$

where \mathbf{E} is the *macroscopic* electric field produced by the polarization in the wave and $\bar{C}_{\alpha\beta}\left(\frac{\mathbf{y}}{kk'}\right)$ is a function that can be expanded in \mathbf{y} . In §4 we shall develop the perturbation theory to set up the lattice wave equation which contains explicitly the electric field.

§3. SEPARATION OF THE MACROSCOPIC ELECTRIC FIELD.

Let

$$\Phi_{\alpha\beta}\left(\frac{l-l'}{kk'}\right) = \Phi_{\alpha\beta}^{(A)}\left(\frac{l-l'}{kk'}\right) + \Phi_{\alpha\beta}^{(\text{Coul})}\left(\frac{l-l'}{kk'}\right), \quad \dots \quad (3.1)$$

where the second term is due to the purely Coulomb interactions and $\Phi_{\alpha\beta}^{(A)}\left(\frac{l-l'}{kk'}\right)$ includes the rest. It is, however, supposed that

$$\Phi_{\alpha\beta}^{(\text{Coul})}\left(\frac{0}{kk}\right) = 0, \quad \dots \quad (3.2)$$

that is, the contribution of the Coulomb restoring force on a particle due to its own displacement is supposed to be absorbed in $\Phi_{\alpha\beta}^{(A)}\left(\frac{0}{kk}\right)$. The

explicit expressions for this part is

$$e_k \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \left(\sum_{l' \neq k'} \frac{e_{k'}}{|\mathbf{x}(l') - \mathbf{x}|} \right)_{\mathbf{x}(k)} \quad . \quad . \quad . \quad (3.3)$$

where the prime over the summation excludes $l'=0$, $k'=0$.

Define

$$\Phi_{\alpha\beta}^{(\text{Coul})} \left(\frac{y}{kk'} \right) = \sum_l \Phi_{\alpha\beta}^{(\text{Coul})} \left(\frac{l}{kk'} \right) \exp \{ -2\pi i \mathbf{y} \cdot \mathbf{x}(l) \}. \quad . \quad . \quad . \quad (3.4)$$

This has a simple interpretation. If an electric dipole $e_k \exp \{ 2\pi i \mathbf{y} \cdot \mathbf{x}(l) \}$ parallel to β -axis is placed at each of the k' -lattice points, $-\Phi_{\alpha\beta}^{(\text{Coul})} \left(\frac{y}{kk'} \right)$ is the Coulomb force on the particle $\left(\frac{0}{k} \right)$. For $k=k'$, the contribution of the dipole at $\left(\frac{0}{k} \right)$ is excluded. All these coefficients can be expressed in terms of the function

$$F_{\alpha\beta} \left(\frac{y}{\mathbf{x}} \right) = \sum_l \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \left(\frac{1}{|\mathbf{x}(l) - \mathbf{x}|} \right) \exp \{ 2\pi i \mathbf{y} \cdot \mathbf{x}(l) \}, \quad . \quad . \quad . \quad (3.5)$$

which is the electric field at \mathbf{x} due to dipoles $\exp \{ 2\pi i \mathbf{y} \cdot \mathbf{x}(l) \}$ parallel to β -axis placed on the lattice points $\mathbf{x}(l)$. Thus

$$-\Phi_{\alpha\beta}^{(\text{Coul})} \left(\frac{y}{kk'} \right) = e_k e_{k'} F_{\alpha\beta} \left(\frac{y}{\mathbf{x}(k) - \mathbf{x}(k')} \right) \quad k \neq k', \quad . \quad . \quad (3.6)$$

$$-\Phi_{\alpha\beta}^{(\text{Coul})} \left(\frac{y}{kk} \right) = e_k^2 \lim_{\mathbf{x} \rightarrow 0} \left\{ F_{\alpha\beta} \left(\frac{y}{\mathbf{x}} \right) \right\}, \quad . \quad . \quad . \quad (3.7)$$

where

$$F_{\alpha\beta}^0 \left(\frac{y}{\mathbf{x}} \right) = F_{\alpha\beta} \left(\frac{y}{\mathbf{x}} \right) - \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \left(\frac{1}{|\mathbf{x}|} \right). \quad . \quad . \quad . \quad (3.8)$$

Owing to the identity

$$\frac{2}{\sqrt{\pi}} \int_0^\infty \exp \{ -a^2 \rho^2 \} d\rho = \frac{1}{a}, \quad . \quad . \quad . \quad (3.9)$$

(3.5) can be written as

$$F_{\alpha\beta} \left(\frac{y}{\mathbf{x}} \right) = \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \int_0^\infty \left[\frac{2}{\sqrt{\pi}} \sum_l \exp \{ -|\mathbf{x}(l) - \mathbf{x}|^2 \rho^2 + 2\pi i \mathbf{y} \cdot (\mathbf{x}(l) - \mathbf{x}) \} \right] \times \exp \{ 2\pi i \mathbf{y} \cdot \mathbf{x} \} d\rho. \quad . \quad . \quad . \quad (3.10)$$

The expression included in the bracket is periodic in the lattice and can be expressed as the following Fourier series :

$$\begin{aligned} \frac{2}{\sqrt{\pi}} \sum_l \exp \{ -|\mathbf{x}(l) - \mathbf{x}|^2 \rho^2 + 2\pi i \mathbf{y} \cdot (\mathbf{x}(l) - \mathbf{x}) \} \\ = \frac{2\pi}{v_a h} \rho^{-3} \exp \left\{ 2\pi i \mathbf{y}(h) \cdot \mathbf{x} - \frac{\pi^2}{\rho^2} |\mathbf{y}(h) + \mathbf{y}|^2 \right\}, \quad (3.11) \end{aligned}$$

where $\mathbf{y}(h) = h_1 \mathbf{b}^1 + h_2 \mathbf{b}^2 + h_3 \mathbf{b}^3$ are reciprocal lattice vectors, $h(h_1, h_2, h_3)$ being integers. The Fourier coefficients have been obtained by the usual methods (cf. Kellermann 1940). (3.11) is the theta-function transformation introduced into crystal theory by Ewald (1912) in connection with optical theory. The series on the two sides of (3.11) are fast convergent respectively for large and small values of the parameter ρ . The integral in (3.10) may be split into two regions in which the two respective forms of the series are used :

$$\begin{aligned} F_{\alpha\beta}(\mathbf{y}) = \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \left\{ \frac{2}{\sqrt{\pi}} \sum_l \int_R^\infty \exp \{ -|\mathbf{x}(l) - \mathbf{x}|^2 \rho^2 + 2\pi i \mathbf{y} \cdot \mathbf{x}(l) \} d\rho \right. \\ \left. + \frac{2\pi}{v_a h} \sum_0^R \exp \left\{ 2\pi i (\mathbf{y}(h) + \mathbf{y}) \cdot \mathbf{x} - \frac{\pi^2}{\rho^2} |\mathbf{y}(h) + \mathbf{y}|^2 \right\} \frac{d\rho}{\rho^3} \right\}. \quad (3.12) \end{aligned}$$

By suitable choice of R fast convergence for both series in (3.12) can be secured. However it is more important that the transformation makes it possible to separate the macroscopic electric field from the lattice forces. Let us define for convenience

$$G(x) = x^{-2} \exp \{ -x^2 \}, \quad H(x) = \frac{2x^{-1}}{\sqrt{\pi}} \int_x^\infty \exp \{ -x^2 \} dx. \quad (3.13)$$

(3.12) can be expressed in terms of these functions as

$$\begin{aligned} F_{\alpha\beta}(\mathbf{y}) = \frac{\partial^2}{\partial x_\alpha \partial x_\beta} \left\{ \frac{1}{\pi v_a} |\mathbf{y}|^{-2} \exp \left\{ 2\pi i \mathbf{y} \cdot \mathbf{x} - \frac{\pi^2}{R^2} |\mathbf{y}|^2 \right\} \right. \\ \left. + R \sum_l H(R |\mathbf{x}(l) - \mathbf{x}|) \exp \{ 2\pi i \mathbf{y} \cdot \mathbf{x}(l) \} \right. \\ \left. + \frac{\pi}{v_a R^2} \sum_h G(\pi R^{-1} |\mathbf{y}(h) + \mathbf{y}|) \exp \{ 2\pi i (\mathbf{y}(h) + \mathbf{y}) \cdot \mathbf{x} \} \right\}. \quad (3.14) \end{aligned}$$

The first term comes from the term $h=0$ which will be kept apart from the rest, for it will be found to contain the macroscopic field. The prime over the second summation excludes $h=0$. $H(x)$ is singular at $x=0$. This occurs only in $F_{\alpha\beta}^0(\mathbf{y})$ for the term $l=0$. But in this case the term combines with the additional $-1/|\mathbf{x}|$ in (3.8) and leads to

$$RH^0(R |\mathbf{x}|) = R \left(H(R |\mathbf{x}|) - \frac{1}{R |\mathbf{x}|} \right) = -\frac{1}{|\mathbf{x}|} \int_0^{R|\mathbf{x}|} \exp \{ -x^2 \} dx, \quad (3.15)$$

which tends to $-R$ as $x \rightarrow 0$. Thus the singularity is removed and, moreover, $F_{\alpha\beta}^0\left(\frac{\mathbf{y}}{\mathbf{x}}\right)$ can be obtained from $F_{\alpha\beta}\left(\frac{\mathbf{y}}{\mathbf{x}}\right)$ by replacing $H(R|\mathbf{x}|)$ by $H^0(R|\mathbf{x}|)$ in the term $l=0$. Carrying out the differentiations in (3.14) one finds

$$\begin{aligned} F_{\alpha\beta}\left(\frac{\mathbf{y}}{\mathbf{x}}\right) = & \left\{ -\frac{4\pi}{v_a}\left(\frac{y_\alpha y_\beta}{|\mathbf{y}|^2}\right) \exp\left\{2\pi i \mathbf{y} \cdot \mathbf{x} - \frac{\pi^2 |\mathbf{y}|^2}{R^2}\right\} \right. \\ & + R^3 \sum_l H_{\alpha\beta}(R|\mathbf{x}(l) - \mathbf{x}|) \exp\{2\pi i \mathbf{y} \cdot \mathbf{x}(l)\} \\ & - \frac{4\pi^3}{v_a} \frac{1}{R^2} \sum_h \Sigma'_h (y_\alpha(h) + y_\alpha)(y_\beta(h) + y_\beta) \\ & \left. \times G(\pi R^{-1}|\mathbf{y}(h) + \mathbf{y}|) \exp\{2\pi i(\mathbf{y}(h) + \mathbf{y}) \cdot \mathbf{x}\} \right\}, \quad (3.16) \end{aligned}$$

where

$$H_{\alpha\beta}(|\mathbf{x}|) = \frac{\partial^2}{\partial x_\alpha \partial x_\beta} H(|\mathbf{x}|). \quad (3.17)$$

All terms in (3.16) can be expanded with respect to \mathbf{y} except the first which may be written as

$$\begin{aligned} & -\frac{4\pi}{v_a}\left(\frac{y_\alpha y_\beta}{|\mathbf{y}|^2}\right) \exp\left\{2\pi i \mathbf{y} \cdot \mathbf{x} - \frac{\pi^2 |\mathbf{y}|^2}{R^2}\right\} \\ & = -\frac{4\pi}{v_a}\left(\frac{y_\alpha y_\beta}{|\mathbf{y}|^2}\right) \left\{ 1 - \frac{\pi^2 |\mathbf{y}|^2}{R^2} + \frac{1}{2} \left(\frac{\pi^2 |\mathbf{y}|^2}{R^2}\right)^2 + \dots \right\} \exp\{2\pi i \mathbf{y} \cdot \mathbf{x}\} \\ & = -\frac{4\pi}{v_a}\left(\frac{y_\alpha y_\beta}{|\mathbf{y}|^2}\right) \exp\{2\pi i \mathbf{y} \cdot \mathbf{x}\} \\ & \quad - \frac{4\pi}{v_a} y_\alpha y_\beta \left[|\mathbf{y}|^{-2} \left(\exp\left\{-\frac{\pi^2 |\mathbf{y}|^2}{R^2}\right\} - 1 \right) \right] \exp\{2\pi i \mathbf{y} \cdot \mathbf{x}\}. \quad (3.18) \end{aligned}$$

Written in this form the second term is regular and can be expressed as a power series in \mathbf{y} . Substituting (3.18) in (3.16) we obtain

$$\begin{aligned} F_{\alpha\beta}\left(\frac{\mathbf{y}}{\mathbf{x}}\right) = & -\frac{4\pi}{v_a}\left(\frac{y_\alpha y_\beta}{|\mathbf{y}|^2}\right) \exp\{2\pi i \mathbf{y} \cdot \mathbf{x}\} \\ & + \left\{ -\frac{4\pi}{v_a} y_\alpha y_\beta \left[|\mathbf{y}|^{-2} \left(\exp\left\{-\frac{\pi^2 |\mathbf{y}|^2}{R^2}\right\} - 1 \right) \right] \exp\{2\pi i \mathbf{y} \cdot \mathbf{x}\} \right. \\ & + R^3 \sum_l H_{\alpha\beta}(R|\mathbf{x}(l) - \mathbf{x}|) \exp\{2\pi i \mathbf{y} \cdot \mathbf{x}(l)\} \\ & - \frac{4\pi^3}{v_a} \frac{1}{R^2} \sum_h \Sigma'_h (y_\alpha(h) + y_\alpha)(y_\beta(h) + y_\beta) \\ & \left. \times G(\pi R^{-1}|\mathbf{y}(h) + \mathbf{y}|) \exp\{2\pi i(\mathbf{y}(h) + \mathbf{y}) \cdot \mathbf{x}\} \right\}. \quad (3.19) \end{aligned}$$

The only irregular term in (3.19)

$$-\frac{4\pi}{v_a} \left(\frac{y_\alpha y_\beta}{|\mathbf{y}|^2} \right) \exp \{2\pi i \mathbf{y} \cdot \mathbf{x}\}$$

is in fact the macroscopic field due to the electric polarization in the dipole lattice if the latter is regarded as a Maxwell dielectric continuum. For a polarization wave in a continuum, the polarization charge density is given in terms of the longitudinal component of the polarization vector by

$$\rho = -\nabla \cdot \mathbf{P} \exp \{2\pi i \mathbf{y} \cdot \mathbf{x}\} = -\nabla \cdot \mathbf{P}_{\text{long}} \exp \{2\pi i \mathbf{y} \cdot \mathbf{x}\}. \quad (3.20)$$

The electric field is the particular solution of

$$\nabla \cdot \mathbf{E} = 4\pi\rho = -4\pi\nabla \cdot \mathbf{P}_{\text{long}} \exp \{2\pi i \mathbf{y} \cdot \mathbf{x}\}, \quad (3.21)$$

given by

$$\mathbf{E} = -4\pi\mathbf{P}_{\text{long}} \exp \{2\pi i \mathbf{y} \cdot \mathbf{x}\}. \quad (3.22)$$

In the above dipole lattice the polarization vector is obviously

$$\frac{1}{v_a} \exp \{2\pi i \mathbf{y} \cdot \mathbf{x}\},$$

along the β -axis. The longitudinal component in vector form is

$$\mathbf{P}_{\text{long}} \exp \{2\pi i \mathbf{y} \cdot \mathbf{x}\} = \frac{y_\beta \mathbf{y}}{|\mathbf{y}|^2} \frac{1}{v_a} \exp \{2\pi i \mathbf{y} \cdot \mathbf{x}\}. \quad (3.23)$$

It follows from (3.23) that α -component of the electric field is given by

$$-\frac{4\pi}{v_a} \frac{y_\alpha y_\beta}{|\mathbf{y}|^2} \exp \{2\pi i \mathbf{y} \cdot \mathbf{x}\}. \quad (3.24)$$

This is identical with the first term in (3.19).

The above result can now be combined with the non-Coulomb contributions. From (2.4), (3.1), (3.6), (3.7), (3.19) it is readily found that

$$\begin{aligned} C_{\alpha\beta} \left(\frac{y}{kk'} \right) &= \frac{1}{\sqrt{(m_k m_{k'})}} \sum_l \Phi_{\alpha\beta} \left(\frac{l}{kk'} \right) \exp \left\{ -2\pi i \mathbf{y} \cdot \mathbf{x} \left(\frac{l}{kk'} \right) \right\} \\ &= \frac{4\pi}{v_a} \left(\frac{y_\alpha y_\beta}{|\mathbf{y}|^2} \right) \frac{e_k e_{k'}}{\sqrt{(m_k m_{k'})}} + \bar{C}_{\alpha\beta} \left(\frac{y}{kk'} \right), \end{aligned} \quad (3.25)$$

where

$$\begin{aligned} \bar{C}_{\alpha\beta} \left(\frac{y}{kk'} \right) &= \frac{1}{\sqrt{(m_k m_{k'})}} \left\{ \sum_l \Phi_{\alpha\beta}^{(A)} \left(\frac{l}{kk'} \right) \exp \left\{ -2\pi i \mathbf{y} \cdot \mathbf{x} \left(\frac{l}{kk'} \right) \right\} \right. \\ &\quad \left. - e_k e_{k'} \left[-\frac{4\pi}{v_a} y_\alpha y_\beta \left(|\mathbf{y}|^{-2} \left[\exp \left\{ \frac{-\pi^2 |\mathbf{y}|^2}{R^2} \right\} - 1 \right] \right) \right. \right. \\ &\quad \left. \left. + R^3 \sum_l H_{\alpha\beta} \left(R \left| \mathbf{x} \left(\frac{l}{kk'} \right) \right| \right) \exp \left\{ 2\pi i \mathbf{y} \cdot \mathbf{x} \left(\frac{l}{kk'} \right) \right\} \right. \right. \\ &\quad \left. \left. - \frac{4\pi^3}{v_a} \frac{1}{R^2} \sum_h (y_\alpha(h) + y_\alpha)(y_\beta(h) + y_\beta) \right. \right. \\ &\quad \left. \left. \times G(\pi R^{-1} |\mathbf{y}(h) + \mathbf{y}|) \exp \left\{ 2\pi i \mathbf{y}(h) \cdot \mathbf{x} \left(\frac{0}{kk'} \right) \right\} \right] \right\}. \end{aligned} \quad (3.26)$$

The equation (2.3) can be written as

$$\omega^2 \begin{pmatrix} y \\ j \end{pmatrix} w_\alpha \begin{pmatrix} k \\ j \end{pmatrix} = - \frac{e_k}{\sqrt{m_k}} E_\alpha + \sum_{k'\beta} \bar{C}_{\alpha\beta} \begin{pmatrix} y \\ kk' \end{pmatrix} w_\beta \begin{pmatrix} k' \\ j \end{pmatrix}, \quad (3.27)$$

where \mathbf{E} is the macroscopic electric field in cell $l=0$ produced by the electric polarization in the lattice wave; explicitly,

$$E_\alpha = - \frac{4\pi}{v_a} \begin{pmatrix} y_\alpha \\ |\mathbf{y}| \end{pmatrix} \sum_{k\beta} \begin{pmatrix} y_\beta \\ |\mathbf{y}| \end{pmatrix} \frac{e_{k'}}{\sqrt{m_{k'}}} w_\beta \begin{pmatrix} k' \\ j \end{pmatrix}. \quad (3.28)$$

(2.3) can be regarded as the equation of vibration for particles in the cell $l=0$ used as representative cell. (3.27) shows that *part of the lattice interaction appears as a superimposed macroscopic electric field*, only the remainder plays the rôle of the ordinary lattice restoring forces.

§ 4. BORN'S PERTURBATION METHOD.

As already mentioned $\bar{C}_{\alpha\beta} \begin{pmatrix} y \\ kk' \end{pmatrix}$ can be expanded in \mathbf{y} , so we can follow Born's procedure to obtain the elastic wave equations. The following general identities will be needed for subsequent discussions:

$$\sum_{l,k'} \Phi_{\alpha\beta} \begin{pmatrix} l \\ kk' \end{pmatrix} = 0, \quad (4.1)$$

$$\sum_{l,k'} \Phi_{\alpha\beta} \begin{pmatrix} l \\ kk' \end{pmatrix} x_\gamma \begin{pmatrix} l \\ kk' \end{pmatrix} = 0, \quad (4.2)$$

$$\sum_{l,k'} \Phi_{\alpha\beta} \begin{pmatrix} l \\ kk' \end{pmatrix} x_\gamma \begin{pmatrix} l \\ kk' \end{pmatrix} = \sum_{l,k'} \Phi_{\alpha\gamma} \begin{pmatrix} l \\ kk' \end{pmatrix} x_\beta \begin{pmatrix} l \\ kk' \end{pmatrix}, \quad (4.3)$$

$$\sum_{l,k'} \Phi_{\alpha\beta} \begin{pmatrix} l \\ kk' \end{pmatrix} x_\gamma \begin{pmatrix} l \\ kk' \end{pmatrix} x_\lambda \begin{pmatrix} l \\ kk' \end{pmatrix} = \sum_{l,k'} \Phi_{\alpha\lambda} \begin{pmatrix} l \\ kk' \end{pmatrix} x_\gamma \begin{pmatrix} l \\ kk' \end{pmatrix} x_\beta \begin{pmatrix} l \\ kk' \end{pmatrix}. \quad (4.4)$$

The proofs for the relations (4.1) and (4.3) can be found in the paper by Begbie and Born (1947). The proofs for (4.2) and (4.4) will be given in a book being prepared by Professor Max Born and the author; to save space they will not be given here. When there are charged particles, the sums in the above identities are conditionally convergent. In such cases an exponential factor $\exp \{-\alpha r\}$ can be introduced in the Coulomb interaction and the sums can be taken as the limiting values for $\alpha \rightarrow 0$ (cf. Ewald 1916).

To solve (3.27) one makes the following formal expansions with respect to ϵ for given \mathbf{y} where ϵ is an order parameter and may be taken as unity in the results.

$$\omega \begin{pmatrix} \epsilon y \\ j \end{pmatrix} = \epsilon \omega^{(1)} \begin{pmatrix} y \\ j \end{pmatrix} + \dots, \quad (4.5)$$

$$w_\alpha \begin{pmatrix} k \\ j \end{pmatrix} = w_\alpha^{(0)} \begin{pmatrix} k \\ j \end{pmatrix} + \epsilon w_\alpha^{(1)} \begin{pmatrix} k \\ j \end{pmatrix} + \frac{1}{2} \epsilon^2 w_\alpha^{(2)} \begin{pmatrix} k \\ j \end{pmatrix} + \dots, \quad (4.6)$$

$$E_\alpha = E_\alpha^{(0)} + \epsilon E_\alpha^{(1)} + \frac{1}{2} \epsilon^2 E_\alpha^{(2)} + \dots, \quad (4.7)$$

$$\bar{C}_{\alpha\beta}\left(\frac{\epsilon y}{kk'}\right) = \bar{C}_{\alpha\beta}^{(0)}(kk') + \epsilon \sum_{\gamma} \bar{C}_{\alpha\beta, \gamma}^{(1)}(kk') y_{\gamma} + \frac{1}{2} \epsilon^2 \sum_{\gamma\lambda} \bar{C}_{\alpha\beta, \gamma\lambda}^{(2)}(kk') y_{\gamma} y_{\lambda} + \dots \quad (4.8)$$

According to (3.28) \mathbf{E} is linear in the displacements, so the different orders $E_{\alpha}^{(0)}$, $E_{\alpha}^{(1)}$, ... in (4.7) correspond to the different order terms of (4.6). The coefficients in (4.8) can be obtained by the straightforward expansion of (3.26). The resulting expressions are somewhat lengthy and need not be written down explicitly. One notices that

$$\begin{aligned} \sum_{k'} \bar{C}_{\alpha\beta}\left(\frac{y}{kk'}\right) \sqrt{m_{k'}} &= \frac{4\pi}{v_a} \left(\frac{y_{\alpha} y_{\beta}}{|\mathbf{y}|^2} \right) \frac{e_k}{\sqrt{m_k}} \sum_{k'} e_{k'} + \sum_{k'} \bar{C}_{\alpha\beta}\left(\frac{y}{kk'}\right) \sqrt{m_{k'}} \\ &= \sum_{k'} \bar{C}_{\alpha\beta}\left(\frac{y}{kk'}\right) \sqrt{m_{k'}}, \quad \dots \quad (4.9) \end{aligned}$$

for

$$\sum_{k'} e_{k'} = 0. \quad \dots \quad (4.10)$$

Therefore it follows from (4.1), (4.2), (4.3), (4.4) that

$$\sum_{k'} \bar{C}_{\alpha\beta}^{(0)}(kk') \sqrt{m_{k'}} = \frac{1}{\sqrt{m_k}} \sum_{kk'} \Phi_{\alpha\beta}\left(\frac{l}{kk'}\right) = 0, \quad \dots \quad (4.11)$$

$$\sum_{kk'} \bar{C}_{\alpha\beta, \gamma}^{(1)}(kk') \sqrt{(m_k m_{k'})} = -2\pi i \sum_{kk'} \Phi_{\alpha\beta}\left(\frac{l}{kk'}\right) x_{\gamma}\left(\frac{l}{kk'}\right) = 0, \quad \dots \quad (4.12)$$

$$\begin{aligned} \sum_{k'} \bar{C}_{\alpha\beta, \gamma}^{(1)}(kk') \sqrt{m_{k'}} &= -\frac{2\pi i}{\sqrt{m_k}} \sum_{kk'} \Phi_{\alpha\beta}\left(\frac{l}{kk'}\right) x_{\gamma}\left(\frac{l}{kk'}\right) \\ &= \sum_{k'} \bar{C}_{\alpha\gamma, \beta}^{(1)}(kk') \sqrt{m_{k'}} = -\sum_{k'} \bar{C}_{\beta\alpha, \gamma}^{(1)}(k'k) \sqrt{m_{k'}}. \quad (4.13) \end{aligned}$$

$$\begin{aligned} \sum_{k'} \bar{C}_{\alpha\beta, \gamma\lambda}^{(2)}(kk') \sqrt{m_{k'}} &= -\frac{4\pi^2}{\sqrt{m_k}} \sum_{kk'} \Phi_{\alpha\beta}\left(\frac{l}{kk'}\right) x_{\gamma}\left(\frac{l}{kk'}\right) x_{\lambda}\left(\frac{l}{kk'}\right) \\ &= \sum_{k'} \bar{C}_{\alpha\lambda, \gamma\beta}^{(2)}(kk') \sqrt{m_{k'}} = \sum_{k'} \bar{C}_{\beta\alpha, \gamma\lambda}^{(2)}(k'k) \sqrt{m_{k'}}. \quad (4.14) \end{aligned}$$

The equations of zeroth order are

$$0 = -\frac{e_k}{\sqrt{m_k}} E_{\alpha}^{(0)} + \sum_{k'\beta} \bar{C}_{\alpha\beta}^{(0)}(kk') w_{\beta}^{(0)}\left(k' \left| \frac{y}{j} \right. \right). \quad \dots \quad (4.15)$$

The acoustic vibrations correspond to the non-trivial solutions

$$w_{\beta}^{(0)}\left(k' \left| \frac{y}{j} \right. \right) = \sqrt{m_{k'}} w_{\beta}^{(0)}(j), \quad \dots \quad (4.16)$$

where $\mathbf{u}^{(0)}(j)$ is an arbitrary vector. (4.16) satisfies (4.15) on account of (4.11) and that

$$\mathbf{E}^{(0)} = 0, \quad \dots \quad (4.17)$$

which follows directly from (4.10) and (3.28).

The equations of first order are

$$0 = -\frac{e_k}{\sqrt{m_k}} E_\alpha^{(1)} + \sum_{k'\beta\gamma} \bar{C}_{\alpha\beta,\gamma}^{(1)}(kk') y_\gamma \sqrt{m_k} u_\beta^{(0)}(j) + \sum_{k'\beta} \bar{C}_{\alpha\beta}^{(0)}(kk') w_\beta^{(1)} \left(k' \left| \begin{matrix} y \\ j \end{matrix} \right. \right). \quad (4.18)$$

The condition for solubility is that the non-homogeneous part should be orthogonal to an arbitrary solution of the form (4.16), and can be obtained by multiplying (4.18) with $\sqrt{m_k}$ and summing over k :

$$0 = -E_\alpha^{(1)} \sum_k e_k + \sum_{\beta\gamma} y_\gamma u_\beta^{(0)}(j) \sum_{kk'} \bar{C}_{\alpha\beta,\gamma}^{(1)}(kk') \sqrt{(m_k m_{k'})}. \quad (4.19)$$

The equation is identically satisfied on account of (4.10) and (4.12). Owing to (4.11) the solution of (4.18) is arbitrary to the addition of a solution of the form (4.16) and the equations determine only the relative base vectors. So one can always suppose that $w_\alpha^{(1)} \left(k \left| \begin{matrix} y \\ j \end{matrix} \right. \right) = 0$ for $k=0$.

In order to obtain such solutions, construct first the reciprocal matrix of the $(3n-3) \times (3n-3)$ matrix $\bar{C}_{\alpha\beta}^{(0)}(kk')$ ($k, k' \neq 0$). Then border the reciprocal matrix so obtained with zeros to form a $3n \times 3n$ matrix $\Gamma_{\alpha\beta}(kk')$, thus

$$\Gamma_{\alpha\beta}(kk') = 0, \quad k \text{ or } k' = 0 \quad (4.20)$$

and

$$\sum_{k''=1}^{n-1} \sum_\gamma \Gamma_{\alpha\beta}(kk'') \bar{C}_{\gamma\beta}^{(0)}(k''k') = \delta_{kk'} \delta_{\alpha\beta}, \quad k, k' \neq 0.$$

The solution of (4.18) is obtained by multiplying (4.18) with $\Gamma_{\alpha\beta}(kk')$

$$w_\alpha^{(1)} \left(k \left| \begin{matrix} y \\ j \end{matrix} \right. \right) = \sum_{k'\beta} \Gamma_{\alpha\beta}(kk') \frac{e_k}{\sqrt{m_k}} E_\beta^{(1)} - \sum_{k'\mu} \Gamma_{\alpha\mu}(kk') \sum_{k''\beta\gamma} \bar{C}_{\mu\beta,\gamma}^{(1)}(k'k'') y_\gamma \sqrt{m_{k'}} u_\beta^{(0)}(j). \quad (4.21)$$

$\bar{C}_{\alpha\beta}^{(0)}(kk')$ is obtained from (3.26) by putting $\mathbf{y}=0$. It is then seen readily that $\bar{C}_{\alpha\beta}^{(0)}(kk')$ is symmetric in the indices $(k\alpha)$ and $(k'\beta)$. Hence

$$\Gamma_{\alpha\beta}(kk') = \Gamma_{\beta\alpha}(k'k). \quad (4.22)$$

The second order equations are

$$\begin{aligned} \left[\omega^{(1)} \left(\begin{matrix} y \\ j \end{matrix} \right) \right]^2 \sqrt{m_k} u_\alpha^{(0)}(j) = & -\frac{e_k}{\sqrt{m_k}} E_\alpha^{(2)} + \frac{1}{2} \sum_{k'\beta\gamma\lambda} \bar{C}_{\alpha\beta,\gamma\lambda}^{(2)}(kk') y_\gamma y_\lambda \sqrt{m_k} u_\beta^{(0)}(j) \\ & + \sum_{k'\mu\gamma} \bar{C}_{\alpha\mu,\gamma}^{(1)}(kk') y_\gamma \sum_{k''\beta} \Gamma_{\mu\beta}(k'k'') \frac{e_{k''}}{\sqrt{m_{k''}}} E_\beta^{(1)} \\ & - \sum_{k'\mu\gamma} \bar{C}_{\alpha\mu,\gamma}^{(1)}(kk') y_\gamma \sum_{k''\nu} \Gamma_{\mu\nu}(k'k'') \sum_{k'''\beta\lambda} \bar{C}_{\nu\beta,\lambda}^{(1)}(k''k''') y_\lambda \sqrt{m_{k''}} u_\beta^{(0)}(j) \\ & + \frac{1}{2} \sum_{k'\beta} \bar{C}_{\alpha\beta}^{(0)}(kk') w_\beta^{(2)} \left(k' \left| \begin{matrix} y \\ j \end{matrix} \right. \right). \quad (4.23) \end{aligned}$$

The condition of solubility is as before obtained by multiplying (4.23) by $\sqrt{m_k}$ and summing over k . On account of (4.10), (4.11), the first and last terms on the right vanish and we may write the condition as

$$\left[\omega^{(1)} \left(\frac{y}{j} \right) \right]^2 \left(\frac{\Sigma m_k}{v_a} \right) u_{\alpha}^{(0)}(j) = 4\pi^2 \Sigma_{\beta} \left(\Sigma_{\gamma\lambda} \{ \alpha\gamma, \beta\lambda \} y_{\gamma} y_{\lambda} \right) u_{\beta}^{(0)}(j) \\ + 2\pi i \Sigma_{\beta} \left(\Sigma_{\gamma} \{ \beta, \alpha\gamma \} y_{\gamma} \right) E_{\beta}^{(1)}, \quad . \quad . \quad (4.24)$$

where

$$\{ \alpha\gamma, \beta\lambda \} = \frac{1}{4\pi^2 v_a} \left\{ \frac{1}{2} \Sigma_{kk'} \bar{C}_{\alpha\beta, \gamma\lambda}^{(2)}(kk') \sqrt{(m_k m_{k'})} \right. \\ \left. + \Sigma_{k''k'''} \Sigma_{\mu\nu} \Gamma_{\mu\nu}(k''k''') \left(\Sigma_k \bar{C}_{\mu\alpha, \gamma}^{(1)}(k''k) \sqrt{m_k} \right) \left(\Sigma_{k'} \bar{C}_{\nu\beta, \lambda}^{(1)}(k'''k') \sqrt{m_{k'}} \right) \right\} \\ . \quad . \quad . \quad (4.25)$$

and

$$\{ \beta, \alpha\gamma \} = \frac{i}{2\pi v_a} \Sigma_{kk' \mu} \Sigma \bar{C}_{\mu\alpha, \gamma}^{(1)}(k'k) \sqrt{m_k} \Sigma_{k''} \Gamma_{\mu\beta}(k'k'') \frac{e_{k''}}{\sqrt{m_{k''}}}. \quad . \quad . \quad (4.26)$$

It follows from the relations (4.13), (4.14) these coefficients can also be expressed directly in terms of the derivatives of the energy function Φ as

$$\{ \alpha\gamma, \beta\lambda \} = \frac{1}{v_a} \left\{ -\frac{1}{2} \Sigma_{lkk'} \Phi_{\alpha\beta} \left(\frac{l}{kk'} \right) x_{\gamma} \left(\frac{l}{kk'} \right) x_{\lambda} \left(\frac{l}{kk'} \right) \right. \\ \left. - \Sigma_{k''k'''} \Sigma_{\mu\nu} \frac{\Gamma_{\mu\nu}(k''k''')}{\sqrt{(m_{k''} m_{k'''})}} \left(\Sigma_{lk} \Phi_{\mu\alpha} \left(\frac{l}{k''k} \right) x_{\gamma} \left(\frac{l}{k''k} \right) \right) \right. \\ \left. \times \left(\Sigma_{lk'} \Phi_{\nu\beta} \left(\frac{l}{k'''k'} \right) x_{\lambda} \left(\frac{l}{k'''k'} \right) \right) \right\}, \quad . \quad . \quad . \quad (4.27)$$

$$\{ \beta, \alpha\gamma \} = \frac{1}{v_a} \Sigma_{kk' \mu} \Phi_{\mu\alpha} \left(\frac{l}{k'k} \right) x_{\gamma} \left(\frac{l}{k'k} \right) \Sigma_{k''} \frac{\Gamma_{\mu\beta}(k'k'')}{\sqrt{(m_{k'} m_{k''})}} e_{k''}. \quad . \quad . \quad (4.28)$$

The following symmetry relations follow from (4.3), (4.4) and (4.22):

$$\{ \alpha\gamma, \beta\lambda \} = \{ \beta\lambda, \alpha\gamma \} = \{ \alpha\gamma, \lambda\beta \}, \quad . \quad . \quad . \quad (4.29)$$

$$\{ \beta, \alpha\gamma \} = \{ \beta, \gamma\alpha \}. \quad . \quad . \quad . \quad (4.30)$$

§ 5. COMPARISON WITH ELASTIC WAVE EQUATIONS OF PIEZOELECTRIC CRYSTALS.

In a general piezoelectric medium, apart from temperature effects, all the relevant coefficients relating to its mechanical and electric properties are embodied in the following equations:

$$P_{\alpha} = \Sigma_{\beta} a_{\alpha\beta} E_{\beta} + \Sigma_{\alpha q} e_{\alpha q} s_q, \quad . \quad . \quad . \quad (5.1)$$

In the lattice wave, the polarization vector is obtained from the formal solution (4.21) of the first order equation, namely,

$$\begin{aligned} P_{\alpha}^{(1)} \begin{pmatrix} y \\ j \end{pmatrix} &= \frac{1}{v_{\alpha k}} \sum \frac{e_k}{\sqrt{m_k}} w_{\alpha}^{(1)} \begin{pmatrix} k \\ j \end{pmatrix} \\ &= \frac{1}{v_{\alpha \beta}} \sum \left(\sum_{kk'} \frac{e_k e_{k'}}{\sqrt{(m_k m_{k'})}} \Gamma_{\alpha\beta}(kk') \right) E_{\beta}^{(1)} \\ &\quad - \frac{1}{v_{\alpha kk'\mu}} \sum \frac{e_k}{\sqrt{m_k}} \Gamma_{\alpha\mu}(kk') \sum_{k''\beta\gamma} \bar{C}_{\mu\beta,\gamma}^{(1)}(k'k'') \sqrt{m_{k''}} y_{\gamma} w_{\beta}^{(0)}(j). \end{aligned} \quad (5.13)$$

Denote the polarization wave in the elastic medium by

$$\bar{P}_{\alpha} \exp \{2\pi i \mathbf{y} \cdot \mathbf{x} - i\omega(y)t\} \quad . \quad . \quad . \quad . \quad (5.14)$$

and substitute in (5.6):

$$\bar{P}_{\alpha} = \sum_{\beta} a_{\alpha\beta} \bar{E}_{\beta} + 2\pi i \sum_{\beta\gamma} e_{\alpha,\beta\gamma} y_{\gamma} \bar{u}_{\beta} \quad . \quad . \quad . \quad . \quad (5.15)$$

Comparing this with (5.13), one directly obtains

$$a_{\alpha,\beta} = \frac{1}{v_{\alpha kk'}} \sum \frac{e_k e_{k'}}{\sqrt{(m_k m_{k'})}} \Gamma_{\alpha\beta}(kk'), \quad . \quad . \quad . \quad . \quad (5.16)$$

$$e_{\alpha,\beta\gamma} = \frac{i}{2\pi v_{\alpha kk'\mu}} \sum \frac{e_k}{\sqrt{m_k}} \Gamma_{\alpha\mu}(kk') \sum_{k''} \bar{C}_{\mu\beta,\gamma}^{(1)}(k'k'') \sqrt{m_{k''}} = \{\alpha, \beta\gamma\}. \quad . \quad . \quad (5.17)$$

After permutation of α, β (5.17) becomes identical with (5.11) showing the consistency of the results.

§6. RESULTS.

Since the relations that underlie the expressions for the various coefficients are rather scattered, they are briefly summarized: the macroscopic coefficients $a_{\alpha\beta}$, $e_{\alpha\beta}$ (or $e_{\alpha,\beta\gamma}$) and $c_{\theta\sigma}$ (or $c_{\alpha\gamma,\beta\lambda}$) are defined by (5.1), (5.2) (or (5.6), (5.7)). Their expressions in the lattice theory are given by (5.16), (5.11) and (5.12) respectively. The explicit expressions for the bracket expressions are given by (4.25) (4.26), or alternatively by (4.27), (4.28). The matrix $\Gamma_{\alpha\beta}(kk')$ is defined in terms of $\bar{C}_{\alpha\beta}^{(0)}(kk')$ by (4.20); $\bar{C}_{\alpha\beta}^{(0)}(kk')$ is obtained from (3.26) by putting $\mathbf{y}=0$. Although (4.27), (4.28) for the bracket expressions are simpler, (4.25), (4.26) have to be used for practical calculations because of the slow convergence of the Coulomb interaction. $\bar{C}_{\alpha\beta}^{(1)}(kk')$, $\bar{C}_{\alpha\beta}^{(2)}(kk')$ involved in these expressions are defined by (4.8) and the explicit forms are easily obtained by straightforward expansion of (3.26). The value of R in actual evaluation of $\bar{C}_{\alpha\beta}^{(1)}(kk')$, $\bar{C}_{\alpha\beta}^{(2)}(kk')$ can be chosen arbitrarily to secure quick convergence for the two lattice sums involved. It is to be remembered that in (3.26), if $k=k'$ the H -function in the term $l=0$ has to be replaced by H^0 defined by (3.15), G, H themselves being defined by (3.13), (3.17). Moreover, one should remember that the non-Coulomb $\Phi_{\alpha\beta}^{(A)} \begin{pmatrix} l \\ kk' \end{pmatrix}$ actually includes the Coulomb contribution (3.2) for $l=0, k=k'$,

§7. CONCLUSION.

It only remains to emphasize the physical aspect of the present investigation. It should be observed that the cause of the usual divergence is not a trivial one. The usual procedure includes all the lattice interactions as responsible for the elastic restoring force. The above investigation shows that in fact part of the interaction corresponds in the macroscopic theory to a superimposed electric field and only the remaining part acts effectively as the elastic restoring force. A completely parallel situation is already found in the investigations of Ewald and Born on crystal optics; there the macroscopic electromagnetic field within the crystals are found to be produced by the lattice particles themselves. Therefore there alone, the expressions of various coefficient given by Born find their justification. Also in the more recent investigations, by Lyddane and Herzfeld (1938), Fröhlich and Mott (1939), Lyddane, Sachs and Teller (1941), the difference between the frequencies of long transverse and longitudinal waves of the optical branch was pointed out and ascribed to the additional electric field in the longitudinal waves. Here again the electric field is nothing more than part of the Coulomb field of the lattice particles. As far as the material contained in a volume element is concerned, this field has to be regarded as a superimposed field, thus only the remainder can appear as the effective lattice forces.

In this paper we have assumed that the electric polarization is due solely to the displacements of ions as rigid structures. This assumption is of course at its best an approximation. The ions will in general be deformed by the displacements so that the net dipoles caused by the displacements will be modified. Furthermore, due to the presence of the electric field the ions will be polarized, further contributing to the electric polarization in the crystal. The consideration of such effects necessitates a full discussion of how one should treat the deformability and polarizability of electron cloud in crystals, which is not feasible here. A full treatment of such effects will be given in the book which we have already mentioned.

The author wishes to express his indebtedness to Professor Max Born, F.R.S., to whom he owes his basic knowledge on the subject.

REFERENCES.

- BEGGIE, G. H., and BORN, M., 1947, *Proc. Roy. Soc. A.*, **188**, 179.
 BORN, M., 1923, *Atomtheorie des Festen Zustandes*, 2nd Edition (Teubner).
 BORN, M., and GÖPPERT-MAYER, M., 1933, *Handb. d. Phys.*, XXIV., Part II. (Springer).
 EWALD, P. P., 1912, *Münchener Diss. Göttingen*; 1916, *Ann. d. Phys.*, **49**, 1.
 FRÖHLICH, H., and MOTT, N. F., 1939, *Proc. Roy. Soc. A.*, **171**, 496.
 KELLERMANN, E. W., 1940, *Phil. Trans. Roy. Soc.*, **238**, 513.
 LYDDANE, R. H., and HERZFELD, K. F., 1938, *Phys. Rev.*, **54**, 846.
 LYDDANE, R. H., SACHS, R. G., and TELLER, E., 1941, *Phys. Rev.*, **59**, 673.

LXXI. *A Note on the Transient Response of an Oscillatory Circuit with Recurrent Discharge.*

By A. M. HARDIE, M.A., B.Sc., A.M.I.E.E.,
Metropolitan-Vickers Electrical Co. Ltd., Manchester*.

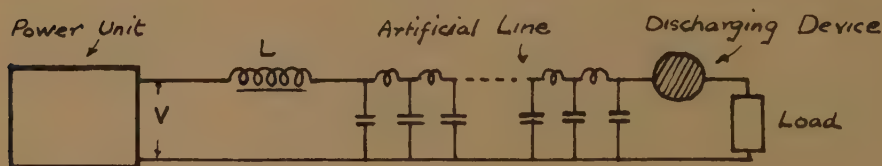
[Received March 23, 1949.]

§ 1. INTRODUCTION.

ONE method of generating short duration high voltage pulses of rectangular shape is by the recurrent charge and discharge of an artificial transmission line (Wilkinson 1946).

The duration of the pulse-forming discharge is generally very much shorter than the charging period, during which the transmission line may be treated as a lumped capacitance. The line is terminated by a load circuit designed to present an impedance equal to the characteristic impedance, while the discharge may be effected by a mercury vapour or hydrogen thyratron, trigatron or rotary spark gap (Craggs, Haine and Meek 1946).

Fig. 1.



Such a system is shown in fig. 1. The line is charged from a source of steady voltage V through an inductance L which may be chosen so that the natural half-period of the oscillatory circuit comprising L and C is greater than the period between successive discharges.

It is well known that with no losses and constant recurrence frequency this system attains a steady state in which the line voltage is $2V$ at the discharge instant.

In the course of an investigation of automatic pulse amplitude control it was necessary to determine the transient behaviour of the system in response to an applied step function of voltage. It was found that the mechanism of the transient possesses interesting features which, as far as is known, have not appeared in the literature.

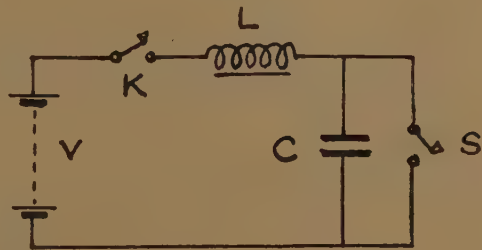
* Communicated by the Author,

§ 2. TRANSIENT RESPONSE.

For the purpose of calculating the transient response the transmission line is replaced by its total capacitance C as in fig. 2. The discharge device is represented by the switch S . The following assumptions are made :

- (a) $\pi\sqrt{LC}>T$, where T is the period between successive discharges, i. e., successive closures of S .
- (b) C is completely discharged at each closure of S in a time which is so short in comparison with T that the discharge may be assumed instantaneous.
- (c) The current in L remains constant during discharge. This follows from (b), and provided L is large, is substantially true in practice.
- (d) Circuit losses are negligible.
- (e) The system is initially quiescent with K and S open.
- (f) At $t=0$, K is closed. At $t=T, 2T, 3T, \dots, NT$, S is closed and opened instantaneously in accordance with (b) above.

Fig. 2.



- Let $\tau=2\pi\sqrt{LC}$,
 $n=1/\sqrt{LC}$,
 $\kappa=\frac{\text{natural period of oscillation}}{\text{period between discharges}}=\frac{\tau}{T}$,
 I =instantaneous current,
 Q =instantaneous charge on C ,
 \bar{I}, \bar{Q} =Laplace Transforms of I and Q ,
 $\overset{\circ}{I}, \overset{\circ}{Q}$ =Initial values of I and Q at the commencement of each charging period.

The differential equations of the circuit are :

$$L\frac{dI}{dt}+\frac{Q}{C}=V, \quad \dots \dots \dots (1)$$

Thus

$$I_{2T} = \frac{V}{nL} \sin nT(1 + \cos nT). \quad . \quad . \quad . \quad (11)$$

During the third charging period, $\dot{I} = I_{2T}$ and $\dot{Q} = 0$ giving

$$I = \frac{V}{nL} [(\sin nt + \sin nT \cos nt + \sin nT \cos nT \cos nt) \quad . \quad . \quad (12)$$

and

$$I_{3T} = \frac{V}{nL} \sin nT(1 + \cos nT + \cos^2 nT). \quad . \quad . \quad . \quad (13)$$

After N charging periods

$$I = \frac{V}{nL} \{ \sin nt + \sin nT(1 + \cos nT + \cos^2 nT + \dots + \cos^{N-2} nT) \cos nt \} \quad . \quad . \quad . \quad (14)$$

and

$$I_{NT} = \frac{V}{nL} \sin nT(1 + \cos nT + \cos^2 nT + \dots + \cos^{N-1} nT). \quad . \quad . \quad (15)$$

On summation of the cosine series and setting $nT = 2\pi/\kappa$, equations (14) and (15) become

$$\begin{aligned} I &= \frac{V}{nL} \left\{ \sin nt + \frac{\sin 2\pi/\kappa}{1 - \cos 2\pi/\kappa} \left(1 - \cos^{N-1} \frac{2\pi}{\kappa} \right) \cos nt \right\}, \\ &= \frac{V}{nL} \left\{ \sin nt + \cot \frac{\pi}{\kappa} \left(1 - \cos^{N-1} \frac{2\pi}{\kappa} \right) \cos nt \right\}, \quad . \quad . \quad (16) \end{aligned}$$

and

$$I_{NT} = \frac{V}{nL} \cot \frac{\pi}{\kappa} \left(1 - \cos^N \frac{2\pi}{\kappa} \right). \quad . \quad . \quad . \quad (17)$$

Equations (16) and (17) are used in plotting the transient current, (17) giving the current at the discharge instants for successive values of N, while (16) gives the intermediate values.

The peak value of the current within each period is obtained by differentiating equation (16) with respect to nt and equating to zero. The maxima occur at

$$nt = \tan^{-1} \left(\frac{\tan \pi/\kappa}{1 - \cos^{N-1} 2\pi/\kappa} \right). \quad . \quad . \quad . \quad (18)$$

When N is so large that the factors containing $\cos^{N-1} 2\pi/\kappa$ and $\cos^N 2\pi/\kappa$ in equations (16) and (17) become sensibly unity, the steady state is

attained. Then

$$\begin{aligned} I &= \frac{V}{nL} \left(\sin nt + \cot \frac{\pi}{\kappa} \cos nt \right), \\ &= \frac{V}{nL} \operatorname{cosec} \frac{\pi}{\kappa} \cos \left(nt - \frac{\pi}{\kappa} \right) \quad . \quad . \quad . \quad (19) \end{aligned}$$

and

$$I_{NT} = \frac{V}{nL} \cot \frac{\pi}{\kappa} \quad . \quad . \quad . \quad (20)$$

Voltage transient.

Equation (4) with $\dot{Q}=0$ becomes

$$\bar{Q} = \frac{\bar{I}}{p},$$

so that

$$\begin{aligned} Q &= \frac{V}{nL} \int_0^t \left\{ \sin nt + \cot \frac{\pi}{\kappa} \left(1 - \cos^{N-1} \frac{2\pi}{\kappa} \right) \cos nt \right\} dt, \\ &= \frac{V}{n^2 L} \left\{ 1 - \cos nt + \cot \frac{\pi}{\kappa} \left(1 - \cos^{N-1} \frac{2\pi}{\kappa} \right) \sin nt \right\} \quad . \quad . \quad . \quad (21) \end{aligned}$$

Thus

$$v = V \left\{ 1 - \cos nt + \cot \frac{\pi}{\kappa} \left(1 - \cos^{N-1} \frac{2\pi}{\kappa} \right) \sin nt \right\} \quad . \quad . \quad . \quad (22)$$

and

$$v_{NT} = V \left\{ 1 - \cos \frac{2\pi}{\kappa} + \cot \frac{\pi}{\kappa} \sin \frac{2\pi}{\kappa} \left(1 - \cos^{N-1} \frac{2\pi}{\kappa} \right) \right\}, \quad . \quad . \quad (23)$$

where v and v_{NT} are respectively the instantaneous voltage and the voltage at the discharge instant.

When the steady state is reached*

$$\begin{aligned} v_{NT} &= V \left(1 - \cos \frac{2\pi}{\kappa} + \cot \frac{\pi}{\kappa} \sin \frac{2\pi}{\kappa} \right) \\ &= 2V. \quad . \quad . \quad . \quad (24) \end{aligned}$$

This is easily shown to be true from energy considerations, but the progress of the transient is not clear except by a step by step analysis.

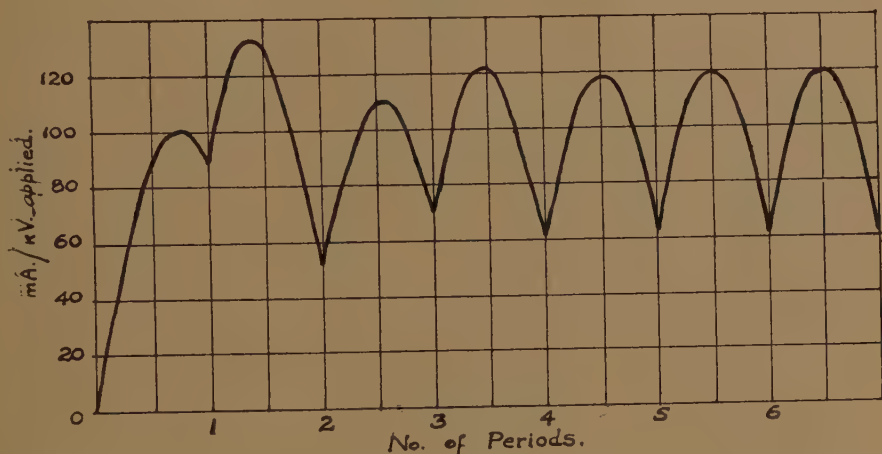
§ 3. GENERAL CHARACTER OF TRANSIENT.

That the current and voltage transients change their character with the value of κ may be seen from an examination of the series expressions for the current and voltage at the discharge instants. The former is

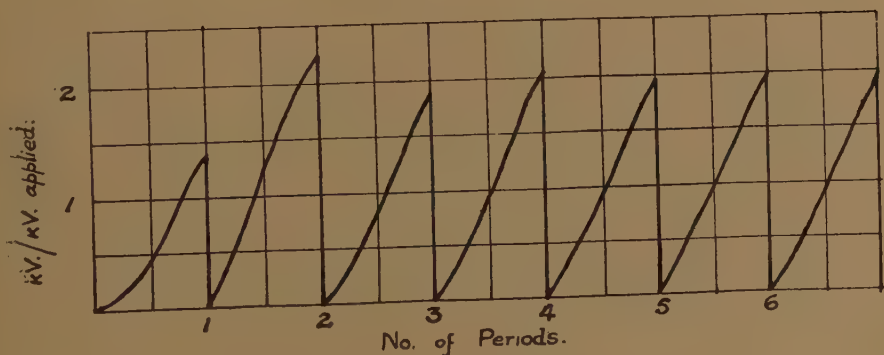
given by equation (15), while the latter is contained in equation (23) which may be rewritten as

$$v_{NT} = V \left\{ 1 - \cos \frac{2\pi}{\kappa} + \sin^2 \frac{2\pi}{\kappa} \left(1 + \cos \frac{2\pi}{\kappa} + \cos^2 \frac{2\pi}{\kappa} + \dots + \cos^{N-2} \frac{2\pi}{\kappa} \right) \right\}. \quad \dots (25)$$

Fig. 3.



(a) Current.



(b) Voltage.

Transient response to voltage-step function.

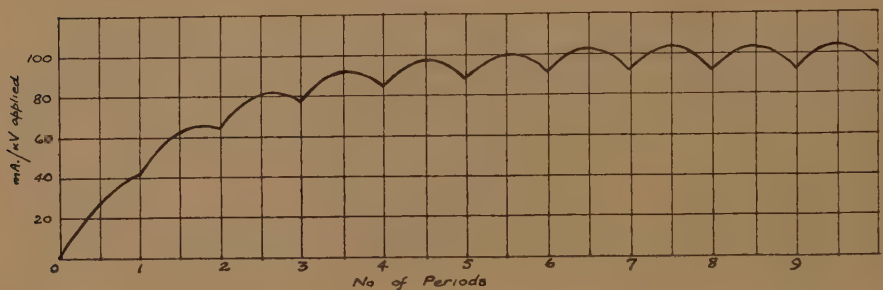
$$L=100 \text{ H}, \quad C=1 \text{ } \mu\text{fd}, \quad \kappa=\pi, \quad T=0.02 \text{ sec.}$$

When $-1 < \cos 2\pi/\kappa < 0$, i. e., $2 < \kappa < 4$, the terms of the cosine series alternate in sign. The sequence formed by the partial sums is oscillatory and approaches the limit $1/(1 - \cos 2\pi/\kappa)$. The transient current and voltage peaks thus oscillate with decreasing amplitude about the steady-state values (fig. 3).

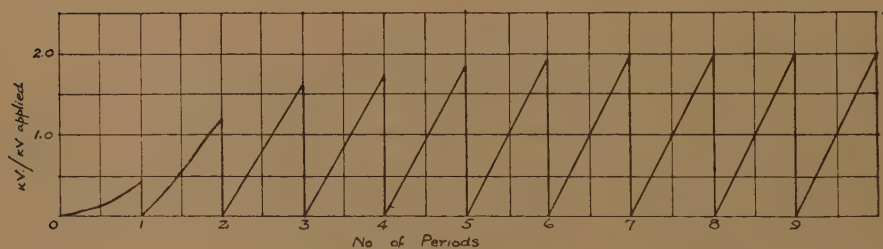
When $\cos 2\pi/\kappa > 0$, i. e., $\kappa > 4$, the terms of the cosine series have the same sign. The sequence of partial sums is monotonic increasing and approaches the limit $1/(1 - \cos 2\pi/\kappa)$ from below. The transient current and voltage peaks then exhibit a smooth rise to the steady state values (fig. 4).

When $\cos 2\pi/\kappa = 0$, i. e., $\kappa = 4$, the transient condition is confined to one charging period only. Then $I_T = I_{2T} = \dots = I_{NT} = V/nL$. The

Fig. 4.



(a) Current.



(b) Voltage.

Transient response to voltage-step function.

$$L=400 \text{ H}, \quad C=1 \text{ } \mu\text{fd}, \quad \kappa=2\pi, \quad T=0.02 \text{ sec.}$$

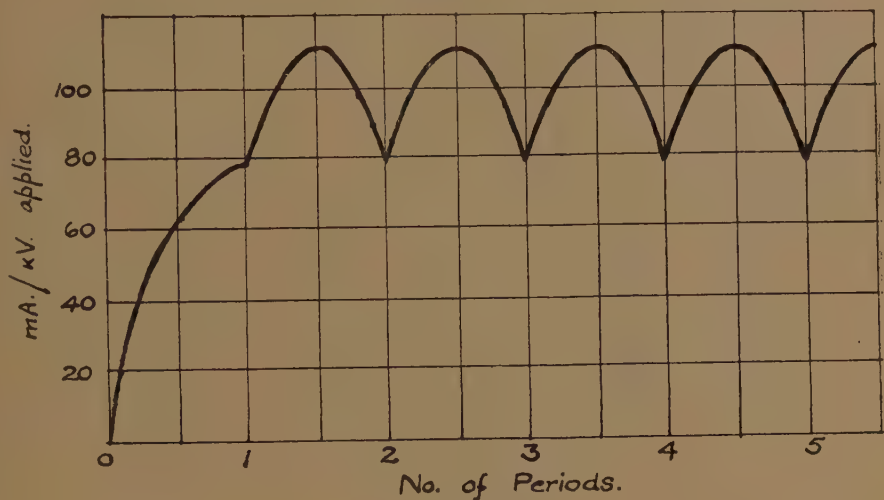
peak current occurs at $nt = \pi/\kappa$ except during the first period and has the value $\sqrt{2}V/nL$, while the voltage reaches $2V$ at the second discharge (fig. 5).

When $\cos 2\pi/\kappa = -1$, i. e., $\kappa = 2$, $I_T = I_{2T} = \dots = I_{NT} = 0$ and the discharge occurs at the instant of zero current, the peak voltage rising to $2V$ at the first discharge. In this case there is no transient condition preceding the steady state, as such. This is, of course, due to the fact that the recurrence frequency of the discharge is equal to twice the natural frequency of the circuit and the current is passing through zero at the

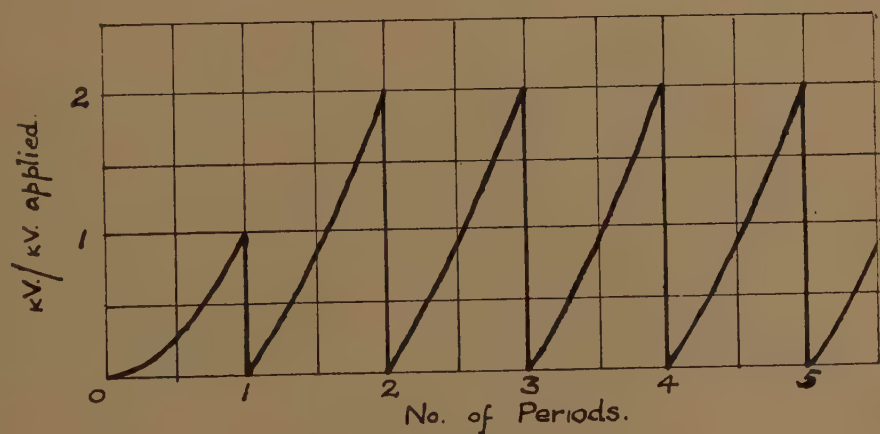
discharge instant. This might be termed a "repeated transient regime" (fig. 6).

The value $\kappa=4$ marks the change from an oscillatory to a non-oscillatory

Fig. 5.



(a) Current.



(b) Voltage.

Transient response to voltage-step function.

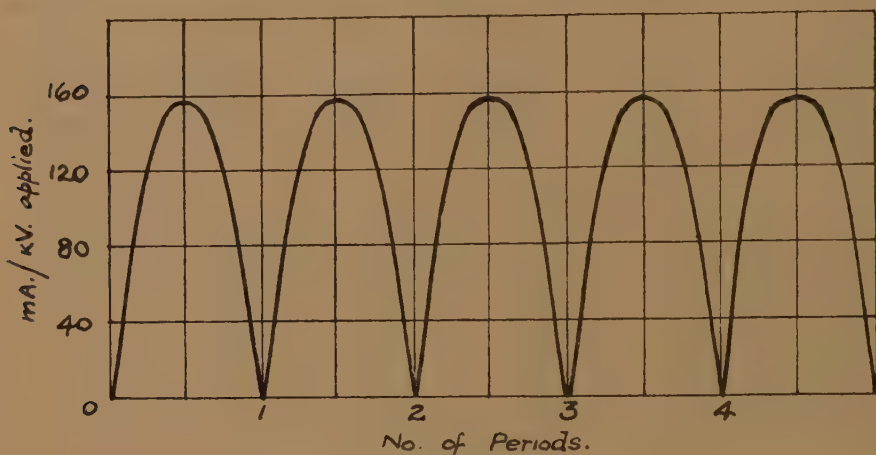
$$L=1600/\pi^2 \text{ H, } C=1 \mu\text{fd, } \kappa=4, T=0.02 \text{ sec.}$$

transient condition. It also affects the position of successive current maxima. These are given by equation (18), i. e.,

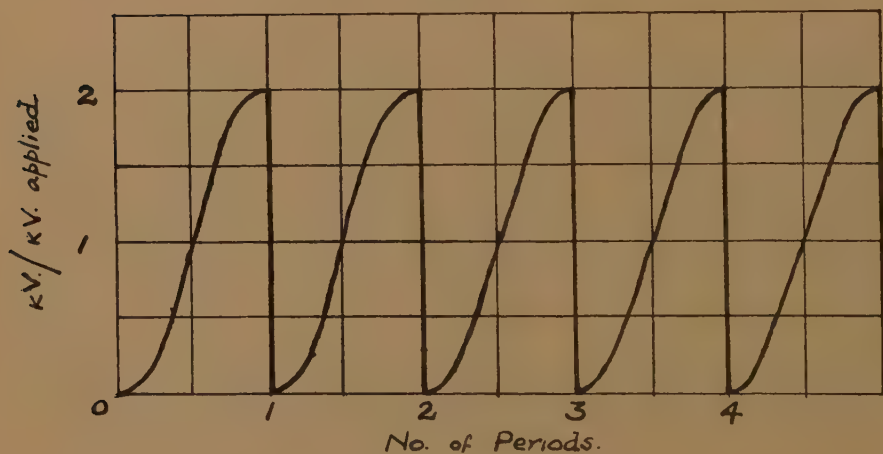
$$nt=\tan^{-1}\left(\frac{\tan \pi/\kappa}{1-\cos^{N-1} 2\pi/\kappa}\right).$$

If $2 < \kappa < 4$, $-1 < \cos 2\pi/\kappa < 0$, then $(1 - \cos^{N-1} 2\pi/\kappa)$ oscillates with increase of N , as does nt . The position of successive maxima therefore oscillates about the half-period point $nt = \pi/\kappa$ with decreasing amplitude.

Fig. 6.



(a) Current.



(b) Voltage.

Transient response to voltage-step function.

$$L = 400/\pi^2 \text{ H}, \quad C = 1 \text{ } \mu\text{fd}, \quad \kappa = 2, \quad T = 0.02 \text{ sec.}$$

If $\kappa > 4$, $\cos 2\pi/\kappa > 0$, so that $(1 - \cos^{N-1} 2\pi/\kappa)$ increases with N while nt decreases. Successive maxima thus approach $nt = \pi/\kappa$ from above.

§ 4. DURATION OF TRANSIENT.

The number of charging periods required for the system to attain a steady state in response to an applied step function of voltage may be determined from equation (16). The steady state is reached when N

is so large that

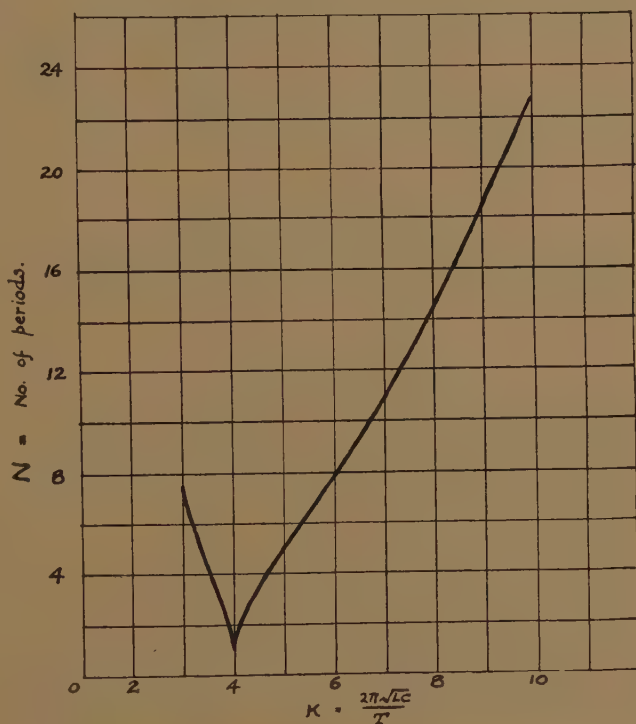
$$\left| \cos^{N-1} \frac{2\pi}{\kappa} \right| < \epsilon, \quad \dots \dots \dots (26)$$

where ϵ is an arbitrary small number.

If $\epsilon=0.01$, then the number of periods required to reach 99 per cent of the steady state values is given by

$$N > \frac{2}{\log_{10} |\cos 2\pi/\kappa|} + 1. \quad \dots \dots \dots (27)$$

Fig. 7.



No. of periods required to attain 99 per cent of steady state values.

$$\text{Graph of } N = \frac{2}{\log_{10} |\cos 2\pi/\kappa|} + 1.$$

Equation (27) is plotted in fig. 7. The curve extends to infinity at $\kappa=2$. This fact supports the conception of a continuously repeated transient as the only steady state.

§ 5. INCLUSION OF CIRCUIT LOSS.

The total circuit loss consists of the energy dissipated in the iron core of the charging inductance, in the resistance of the winding and in the capacitance. The latter may be neglected.

It may be shown that in the steady state the current ripple, expressed as a fraction of the peak current, is given by

$$\frac{\tan \pi/2\kappa}{\operatorname{cosec} \pi/\kappa},$$

so that the iron loss is smaller when κ is increased. During the transient condition, however, the energy loss per period is not constant and an exact analysis has not been undertaken.

The total loss is approximated by a resistance placed in series with the charging inductance. The transient response has been calculated and it is found that, although the resulting expressions are cumbersome due to the presence of exponential terms, the transient mechanism remains the same except in details of amplitude and the rate of reaching the steady state.

The equations for the transient current and the steady state voltage are given in an Appendix.

ACKNOWLEDGMENT.

The author expresses thanks to Sir Arthur P. M. Fleming, and to Mr. B. G. Churcher, for permission to publish this paper.

APPENDIX

Calculation of Transient Response when Resistive Loss is included.

If a resistance R is included in series with L (fig. 2) the transform equations (3) and (4) become

$$Lp\bar{I} + R\bar{I} + \frac{\bar{Q}}{C} = \frac{V}{p} + L\dot{\bar{I}}, \quad \dots \dots \dots (28)$$

$$p\bar{Q} = \bar{I} + \dot{\bar{Q}}. \quad \dots \dots \dots (29)$$

A similar period by period calculation then yields the current during the N th period,

$$I = \frac{V}{nL} \exp(-\alpha t) \left[\sin nt + \left\{ \exp(-\alpha T) \sin \frac{2\pi}{\kappa} \cdot \frac{1 - \lambda^{N-1} \exp(-(N-1)\alpha T)}{1 - \lambda \exp(-\alpha T)} \right\} \times \left(\cos nt - \frac{\alpha}{n} \sin nt \right) \right], \quad \dots \dots \dots (30)$$

where $\alpha = \frac{R}{2L},$

$$\lambda = \cos \frac{2\pi}{\kappa} - \frac{\alpha}{n} \sin \frac{2\pi}{\kappa},$$

$$n = \sqrt{\left(\frac{1}{LC} - \frac{R^2}{4L^2} \right)}.$$

The steady state voltage becomes

$$v_{NT} = V \left\{ 1 - \exp(-\alpha T) \left(\cos \frac{2\pi}{\kappa} + \frac{\alpha}{n} \sin \frac{2\pi}{\kappa} \right) + \frac{1}{n^2 LC} \cdot \frac{\exp(-\alpha T) \sin^2 2\pi/\kappa}{1 - \lambda \exp(-\alpha T)} \right\} \quad (31)$$

When $\alpha=0$, equation (31) reduces to $v_{NT}=2V$.

When $\kappa=2$, $v_{NT}=V(1+\exp(-\alpha\pi/n))$.

REFERENCES

- Craggs, J. D., Haine, M. E., MEEK, J. M., 1946, *J.I.E.E.*, **93**, Pt. IIIA, 963.
 CARSLAW, H. S., and JAEGER, J. C., 1948, *Operational Methods in Applied Mathematics*, 2nd Ed. (Oxford).
 WILKINSON, K. J. R., 1946, *J.I.E.E.*, **93**, Pt. IIIA, 1090.

LXXII. The Loss of Energy of Slow Negative Mesons in Matter.

By R. L. ROSENBERG,
 Natal University College, South Africa*.

[Received May 2, 1949.]

§ 1. INTRODUCTION.

THE rate of loss of energy of negative mesons in matter has been calculated by Fermi and Teller (1947) on the assumption that the electrons can be regarded as free. They found that the time taken for the meson to reduce from 2000 to 0 eV. was of the order of 10^{-14} seconds for solids, and 10^{-11} for gases. As this problem is of some importance with regard to the interpretation of the experiments of Conversi, Pancini and Piccioni (1947), it was felt advisable, in the case of non metals, to make a more detailed calculation of this rate of loss for velocities of the meson less than the orbital velocity of the electron. The method used is due to Mott (1931), and is particularly suited for slow-moving heavy particles. This method has been applied only once by Frame (1937) for calculating the effective cross-section for excitation by heavy positive particles. In this paper we shall only be concerned with the loss of energy due to ionization.

If R be the distance of separation between the atomic nucleus and the meson, regarded as fixed, and $\psi_0(R)$ be the electron wave function corresponding to the initial state of this two-centre problem and $\psi_n(R)$ the electron wave function corresponding to the final (in our case, ionized) state, then if V be the interaction potential between electron and meson, Mott defines

$$M_{0n} = \int \psi_0 \frac{\partial V}{\partial z} \psi_n^* d\tau,$$

* Communicated by the Author.

where z is the coordinate of the meson measured along its supposedly straight path and $d\tau$ refers to the electron coordinates. If W_0 and W_n be the energy levels corresponding to ψ_0 and ψ_n and v the velocity of the meson then the probability $|a_n|^2$ of the transition from the state ψ_0 to the state ψ_n is given by

$$a_n = \int \frac{M_{0n}}{W_0 - W_n} \exp(i(W_0 - W_n)z/\hbar v) dz.$$

Although this method clearly indicates the nature of the collision process for small velocities the computations involved are laborious because of the mathematical difficulties in finding ψ_0 , ψ_n , W_0 , W_n which all belong to a two-centre problem. Over-simplification of these functions will reduce the method to a "Born approximation" which we wish to avoid.

§ 1. THE EVALUATION OF ψ_0 AND W_0 .

Throughout the unit of length will be the Bohr radius (ρ) and the unit of energy the Rydberg (R_y).

TABLE I.

R	α	β	$1/\sqrt{(\pi R)C}$	$-W_0$
1	0.880	0.810	67.2	0.00016
1.5	0.962	0.575	4.81	0.053
2	1.000 *	0.357	1.96	0.170
3	1.000 *	0.192	1.13	0.377
4	—	—	—	0.515
5	—	—	—	0.610
6	—	—	—	0.680

* α is approximately 1 for all values of R and for simplicity this approximation was made for large R .

Let the direction of R be taken as the polar line of the spherical coordinates r, ϑ, ϕ of the electron with the nucleus as centre; r_2 is the distance between electron and meson and let the nucleus exert a Coulomb field due to a charge \bar{Z} . We assume $\psi_0 = C \exp(-\alpha r) \exp(\beta r_2)$ where α and β are functions of R to be determined by the method of variations, and C is a normalizing factor. This is best done in ellipsoidal coordinates ξ, η, ϕ (Condon and Morse 1929) and we find for the energy

$$W_0 = \int \psi_0 H \psi_0^* d\tau$$

$$= - \frac{(2\delta - \beta_1) - (2\delta/\beta_1 R - 2\gamma/\alpha_1 R + \alpha_1 - 2\gamma) \tanh \beta_1 R}{1/\beta_1 + (1/\alpha_1 + 1/\alpha_1^2 R - 1/\beta_1^2 R) \tanh \beta_1 R},$$

where $\delta = \bar{Z} + 1$, $\gamma = \bar{Z} - 1$, $\alpha = \frac{1}{2}(\alpha_1 + \beta_1)$, $\beta = \frac{1}{2}(\beta_1 - \alpha_1)$. The values of α and β for which W_0 is a minimum were then found for the case $\bar{Z} = 1$ and are given in Table I. (together with the values of $1/\sqrt{(\pi R)C}$ and $-W_0$).

A critical value R_0 (which can be estimated to be between $\frac{1}{2}$ and 1) exists below which there is no bound solution for the electron. The implications of this fact on the result will be discussed below. For the present contributions of the region $R < R_0$ are being neglected.

The function ψ_n was taken less exact as that for a free particle :

$$\psi_n = C_1 \frac{1}{\sqrt{(\kappa r)}} J_{l+\frac{1}{2}}(\kappa r) P_l^m(\cos \theta) \exp(im\phi),$$

with normalization over a small interval $d\kappa$ giving

$$C_1 = \frac{\kappa}{2} \sqrt{\left(\frac{(2l+1)(l-m)!}{(l+m)!} \right)} \frac{1}{\sqrt{(2\pi)}},$$

where κ^2 is the energy of the ejected electron.

§ 2. THE EVALUATION OF M_{0n} .

Let the perpendicular from the nucleus to the path of the meson be Δ and let the origin of z be at the foot of this perpendicular. If λ be the angle between the directions of R and z , then it follows (Frame 1937) that

$$r_2 \frac{\partial r_2}{\partial z} = z - r \cos \lambda \cos \theta - r \sin \lambda \sin \theta \cos \phi,$$

and with $V = e^2/r_2$ we have

$$\frac{\partial V}{\partial z} = -\frac{e^2}{\rho^2 r_2^3} \left\{ \frac{z}{R} (R - r \cos \theta) - \frac{\Delta}{R} r \sin \theta \cos \phi \right\}$$

Hence integrating over ϕ we have

$$M_{0n}^0 = -2\pi c c_1 \frac{e^2}{\rho^2} \cdot \frac{z}{R} \int \int \exp(-\alpha r) \exp(\beta r_2) \cdot \frac{1}{r_2^3} (R - r \cos \theta) \frac{1}{\sqrt{(\kappa r)}} \\ \times J_{l+\frac{1}{2}}(\kappa r) P_l^m(\cos \theta) r^2 \sin \theta d\theta dr$$

for the transition to the state κ^2 , l , $m=0$ and

$$M_{0n}^1 = \pi c c_1 \frac{e^2}{\rho^2} \cdot \frac{\Delta}{R} \int \int \exp(-\alpha r) \exp(\beta r_2) \frac{1}{r_2^3} \cdot r \sin \theta \frac{1}{\sqrt{(\kappa r)}} \\ \times J_{l+\frac{1}{2}}(\kappa r) P_l^1(\cos \theta) r^2 \sin \theta d\theta dr$$

for the transition to the states κ^2 , l , $m=\pm 1$.

In the case dealt with by Frame he found it necessary to perform the double integration numerically, but in the above case it is possible to reduce the double integral to a single one. The details of the reduction are given in the appendix. We find

$$M_{0n}^0 = -\frac{e^2}{\rho^2} \cdot \frac{2z}{R} \cdot \frac{c}{\sqrt{(2l+1)}} \cdot \frac{\alpha}{\kappa} \left[\frac{1}{\pi} \int_0^\infty \sqrt{\left(\frac{\pi}{2xR} \right)} g(xR) \frac{dQ_l}{dX} \right. \\ \left. \times \left(\frac{\beta\pi}{2x^2} + \frac{\beta \tan^{-1} \beta/x}{x^2} + \frac{1}{x} \right) dx + (i)^l \int_0^\beta \sqrt{\left(\frac{\pi}{2xR} \right)} g^*(xR) \frac{dQ_l}{dX_1} \cdot \frac{1}{x^2} dx \right]$$

for the case $m=0$, where

$$g(xR) = lJ_{l-\frac{1}{2}}(xR) - (l+1)J_{l+\frac{3}{2}}(xR), \quad X = \frac{\alpha^2 + \kappa^2 + x^2}{2\kappa x},$$

$$g^*(xR) = lI_{l-\frac{1}{2}}(xR) + (l+1)I_{l+\frac{3}{2}}(xR), \quad X_1 = \frac{\alpha^2 + \kappa^2 - x^2}{2i\kappa x}$$

and Q_l is the Legendre function of the second kind; and

$$M_{0n}^1 = -\frac{e^2}{\rho^2} \cdot \frac{\Delta}{R} \cdot \sqrt{\left(\frac{l(l+1)}{2l+1}\right)} \cdot \frac{\alpha}{\kappa} \left[\frac{1}{\pi} \int_0^\infty \sqrt{\left(\frac{\pi}{2xR}\right)} \bar{g}(xR) \frac{dQ_l}{dX} \right. \\ \left. \times \left(\frac{\beta\pi}{2x^2} + \frac{\beta \tan^{-1} \beta/x}{x^2} + \frac{1}{x} \right) dx \right. \\ \left. + (i)^l \int_0^\beta \sqrt{\left(\frac{\pi}{2xR}\right)} \bar{g}^*(xR) \frac{dQ_l}{dX_1} \cdot \frac{1}{x^2} dx \right]$$

for the cases $m=\pm 1$, where

$$\bar{g}(xR) = J_{l-\frac{1}{2}}(xR) + J_{l+\frac{3}{2}}(xR),$$

$$\bar{g}^* = I_{l-\frac{1}{2}}(xR) - I_{l+\frac{3}{2}}(xR).$$

These integrals were computed numerically as functions of R and thus finally

$$a_n = \frac{2\rho^2}{e^2} \int \frac{M_{0n}^0}{W_0 - \kappa^2} \sin\left(\frac{W_0 - \kappa^2}{2v} z\right) dz \quad \text{for } m=0, \\ = \frac{2\rho^2}{e^2} \int \frac{M_{0n}^1}{W_0 - \kappa^2} \cos\left(\frac{W_0 - \kappa^2}{2v} z\right) dz \quad \text{for } m=\pm 1,$$

where v is given in units of $\sqrt{(2R_e/m)}$, i. e. the velocity of the electron in the Bohr orbit.

§ 3. THE CROSS-SECTION.

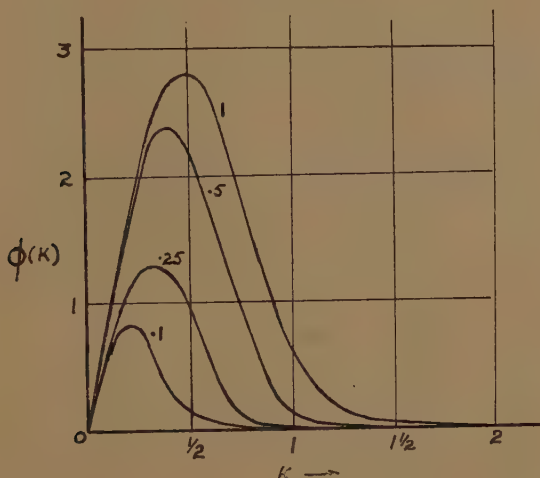
We thus obtain $|a_n|^2$ as a function of Δ , κ , v for different values of l . The convergence in l is rapid, the main contribution being from $l=0$; $l=2$ being negligible. It will be noticed that $M_{0n}^1=0$ for $l=0$ in accordance with the usual transition rules.

The $|a_n|^2$ were then summed for all l and m and after multiplication by $2\pi\Delta d\Delta$ were integrated over Δ . This gives the cross-section $\phi(\kappa)$ for ionization as a function of κ and v , which is shown in fig. 1 (the cross-section in units of ρ^2). Final integration over κ gives the total cross-section as a function of v as shown in fig. 2.

It will be seen from fig. 1 that the electron is ejected with small energy, there being a sharp maximum for $\kappa^2 < \frac{1}{4}$ (i. e. about 3 eV.), and the cross-section being negligible for $\kappa^2 > 1$ (i. e. 13 eV.). Hence as long as the meson energy is above 20 eV. we need not be concerned about the upper limit in the integration over κ in order to fulfil the conservation of energy. If the

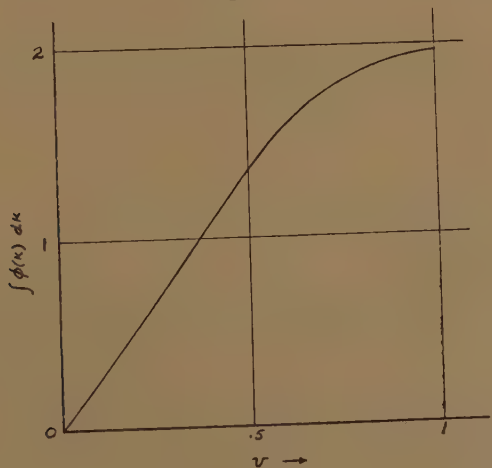
energy of the meson is less than about 20 eV. (*i. e.* $v < 1/10$) the loss of energy in a collision (always greater than 13 eV.) is of the order of its own energy and hence the whole method adopted here becomes invalid in any case. It could be inferred that when $v < 1/10$ the meson is most likely to be captured, whereas for $v > 1/10$ it is unlikely that the meson will be

Fig. 1.



The cross-section as a function of κ . The velocity (v) is indicated on each curve.

Fig. 2.

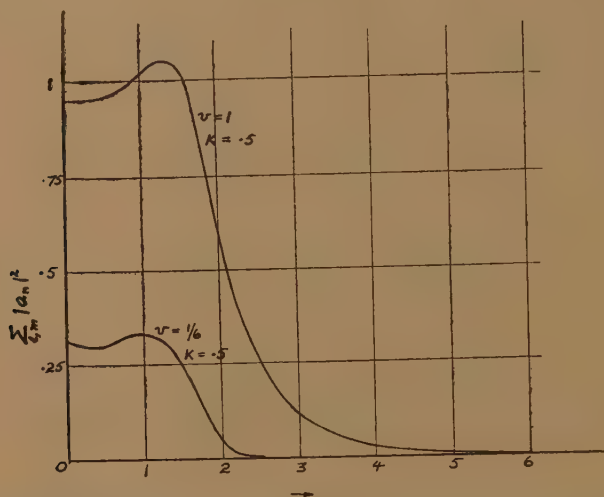


The total cross-section as a function of the velocity.

captured by ionization. If this inference be correct then it may be further inferred from fig. 1 that when captured the orbit of the meson will not have an energy numerically greater than 13 eV., and as the meson mass is about $200m$ it will be captured into an outer orbit provided always that capture is by ionization only without radiative loss by the meson.

Since the higher values of l are negligible we would not expect any great change in angular momentum of the meson when captured, and if $2\pi\Delta |a_n|^2$ be considered as a function of Δ we find a sharp maximum for $\Delta \approx 1.5$, indicating again an outer orbit for the captured meson. (Examples of $|a_n|^2$ as a function of Δ are given in fig. 3.)

Fig. 3.



The total collision probability as a function of the impact parameter Δ , for $\kappa = \frac{1}{2}$.

§ 4. THE RATE OF LOSS OF ENERGY.

The rate of loss of energy dT/dx from $v=1$ downwards will be given by

$$\begin{aligned} -\frac{dT}{dx} &= N \int (1+\kappa^2) \phi(\kappa) d\kappa \\ &= N \cdot f(v), \end{aligned}$$

where N is the number of atoms per unit volume.

$f(v)$ is shown in fig. 4.

For $\Delta < R$ contributions to $f(v)$ have been neglected. This can only decrease the calculated rate of loss of energy, and hence gives the time calculated below the meaning of an upper limit which is just what we intend to obtain.

With $T = \frac{1}{2} M v^2$, where M is the meson mass, we get

$$-M \frac{dv}{dt} = N f(v).$$

Integrating we find

$$\begin{aligned} t &= \frac{M}{N} \int_{v_{\min}}^1 \frac{dv}{f(v)} \rho \sqrt{\left(\frac{2m}{R_y}\right)} \text{ seconds} \\ &= 2.5 \times 10^8 \cdot \frac{M}{N} \text{ seconds,} \end{aligned}$$

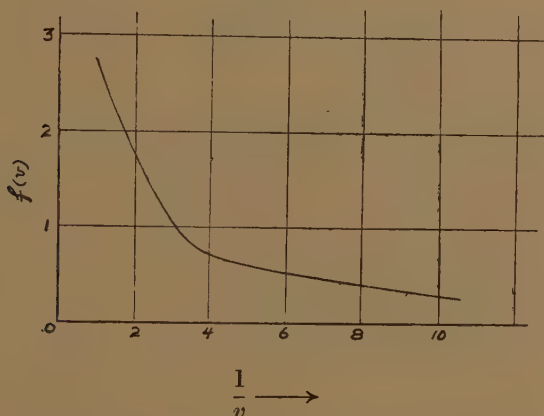
where M is in electron masses and v_{\min} is taken as $1/10$.

With $M=200$ we find in the case of a gas $t=1.8 \times 10^{-9}$ secs., and a range of approximately 2 mm.

It will not be valid to extend the lower limit of the above integral (v_{\min}) any further for reasons given in § 3. All the above calculations are based on the case where the effective nuclear charge $\bar{Z}=1$. Increasing \bar{Z} will diminish $|a_n|^2$ and hence increase the time t , but as we are concerned mainly with the ionization of the outer electrons the value of \bar{Z} will not be very much different from unity. For greater \bar{Z} than 1 the energy loss in a collision will be greater and the form adopted for ψ_n will be even less exact. The validity of the above method for small v will be more doubtful and the cut-off at v_{\min} will have to be higher.

To apply the above results to non conducting solids we should bear in mind (1) that the electron is not so strongly bound as in a Coulomb field with $\bar{Z}=1$, (2) that the energy levels W_0 and W_n are closer. The effect

Fig. 4.



The rate of loss of energy as a function of the reciprocal of the velocity.

of (1) will be to increase $|a_n|^2$ and thus diminish t , whereas (2) will, in the main, mean that the meson loses its energy in smaller amounts and thus increase t . These two effects will therefore counteract each other to a certain extent and may not change the order of magnitude of t (provided of course we use the N value corresponding to a solid). Further, (2) means that we can reduce the lower limit v_{\min} which would increase the time t . There is, however, not much to be gained by doing so for the rate of loss of energy will be overshadowed by the loss of energy due to the interaction of the meson and the lattice vibrations in this region of small velocities (Fröhlich 1937, Fröhlich and Pelzer 1948).

The effect of the screening of the meson charge will not introduce a change of order of magnitude as is apparent from the rapid decrease of $|a_n|^2$ with Δ as shown in fig. 3. In the final integration over z in § 2 it was found that the contributions from large z values were relatively small.

The result found above could be used therefore to give the order of magnitude of t also for solids. Using the appropriate values of N we find times of the order of 10^{-12} seconds and ranges of about 1μ .

I would like to thank Professor N. F. Mott for allowing me to spend a session in his department, and Professor H. Fröhlich for many discussions and help in the preparation of this paper.

REFERENCES.

- CONDON and MORSE, 1929, *Quantum Mechanics* (New York: McGraw-Hill.)
 CONVERSI, PANCINI and PICCIONI, 1947, *Phys. Rev.*, **71**, 209.
 FERMI and TELLER, 1947, *Phys. Rev.*, **72**, 399.
 FERMI, TELLER and WEISSKOPF, 1947, *Phys. Rev.*, **71**, 314.
 FRAME, 1937, *Proc. Camb. Phil. Soc.*,
 FRÖHLICH, 1937, *Proc. Roy. Soc. A.*, **160**, 230.
 FRÖHLICH and PELZER, 1948, *E.R.A. Report*, L/T **184**,
 MOTT, 1931, *Proc. Camb. Phil. Soc.*, **27**, 553.

APPENDIX.

The evaluation of the integral

$$I = \iint \exp(-\alpha r) \exp(\beta r_2) (R - r \cos \theta) \frac{1}{r_2^3} \frac{1}{\sqrt{\kappa r}} \\ \times J_{l+\frac{1}{2}}(\kappa r) P_l(\cos \theta) r^2 \sin \theta d\theta dr.$$

We have (Watson 1944 a)

$$\frac{\exp(\beta r_2)}{r_2} = \pi \Sigma (-)^m (m + \frac{1}{2}) f_m(\beta, r, R) P_m(\cos \theta), \quad \dots (1)$$

where

$$f_m(\beta, r, R) = \frac{1}{\sqrt{(rR)}} (I_{m+\frac{1}{2}}(\beta r) I_{m+\frac{1}{2}}(\beta R) + I_{m+\frac{1}{2}}(\beta r) I_{-m-\frac{1}{2}}(\beta R) \\ = \frac{1}{\sqrt{(rR)}} \left((-)^m \frac{2}{\pi} I_{m+\frac{1}{2}}(\beta r) K_{m+\frac{1}{2}}(\beta R) + 2 I_{m+\frac{1}{2}}(\beta r) I_{m+\frac{1}{2}}(\beta R) \right), \\ \dots (2)$$

for $r < R$, and the same expression with r and R interchanged for $r > R$. Then differentiating with respect to R

$$\exp(\beta r_2) \left(\frac{\beta}{r_2} - \frac{1}{r_2^3} \right) (R - r \cos \theta) = \pi \Sigma (-)^m (m + \frac{1}{2}) \frac{\partial f_m}{\partial R} P_m(\cos \theta),$$

and integrating with respect to β

$$\frac{\beta \exp(\beta r_2)}{r_2^2} (R - r \cos \theta) = \pi \Sigma (-)^m (m + \frac{1}{2}) \beta \left\{ \int_{-\infty}^{\beta} f_m d\beta \right\} P_m(\cos \theta) (R - r \cos \theta).$$

Hence by subtraction

$$\frac{\exp(\beta r_2)}{r_2^3} (R - r \cos \theta) = \pi \Sigma (-)^m (m + \frac{1}{2}) \left\{ R \beta \int_{-\infty}^{\beta} f_m d\beta - \frac{\partial f_m}{\partial R} \right\} P_m(\cos \theta) \\ - \pi \Sigma (-)^m (m + \frac{1}{2}) \beta \left\{ \int_{-\infty}^{\beta} f_m d\beta \right\} P_m(\cos \theta) r \cos \theta.$$

Substituting this in the integral and integrating over θ we obtain

$$I = \pi (-)^l \beta \int_0^{\infty} \int_{-\infty}^{\beta} \exp(-\alpha r) \frac{r^2}{\sqrt{(\kappa r)}} J_{l+\frac{1}{2}}(\kappa r) \left\{ R f_l + r \frac{l+1}{2l+1} f_{l+1} \right. \\ \left. + r \frac{l}{2l+1} f_{l-1} \right\} d\beta dr - \pi (-)^l \int_0^{\infty} \exp(-\alpha r) \frac{r^2}{\sqrt{(\kappa r)}} J_{l+\frac{1}{2}}(\kappa r) \frac{\partial f_l}{\partial R} dr. \quad (3)$$

Now, by the property of the K function, for negative arguments we have

$$\int_{-\infty}^{\beta} f_l d\beta = \int_0^{\beta} f_l d\beta + \frac{1}{\sqrt{(rR)}} \int_0^{\infty} (-)^l \cdot \frac{2}{\pi} I_{l+\frac{1}{2}}(\beta r) K_{l+\frac{1}{2}}(\beta R) d\beta. \quad (4)$$

Now the second part of (2) is symmetrical in r and R so that the integration over r can be conducted in one piece from 0 to ∞ . To obtain a symmetrical expression for the first part we use the relation (Watson 1944 b)

$$\int_0^{\infty} \frac{x}{x^2 + \beta^2} J_{\nu}(rx) J_{\nu}(Rx) dx = I_{\nu}(r\beta) K_{\nu}(R\beta), \quad r < R, \quad \beta > 0, \\ = I_{\nu}(R\beta) K_{\nu}(r\beta), \quad r > R, \quad \beta > 0.$$

Hence for all r we have

$$\int_{-\infty}^{\beta} f_l d\beta = \int_0^{\infty} \int_{(\beta \infty)} (-)^l \cdot \frac{2}{\pi} \cdot \frac{1}{\sqrt{(rR)}} J_{l+\frac{1}{2}}(xr) J_{l+\frac{1}{2}}(xR) \frac{x}{x^2 + \beta^2} d\beta dx \\ + \int_0^{\beta} \frac{2}{\sqrt{(rR)}} I_{l+\frac{1}{2}}(\beta r) I_{l+\frac{1}{2}}(\beta R) d\beta, \quad (5)$$

where $\int_{(\beta \infty)}$ means the sum of the integrals from 0 to ∞ and from 0 to β .

Similarly we find for $\partial f_l / \partial R$ valid for all r :

$$-\int_0^{\infty} (-)^l \frac{1}{\sqrt{(rR)}} \left\{ \frac{1}{2R} \cdot \frac{2}{\pi} J_{l+\frac{1}{2}}(xr) J_{l+\frac{1}{2}}(xR) - \frac{2}{\pi} x J_{l+\frac{1}{2}}(xr) J'_{l+\frac{1}{2}}(xR) \right\} \frac{x}{x^2 + \beta^2} dx \\ - \frac{2}{\sqrt{(rR)}} \left[I_{l+\frac{1}{2}}(\beta r) \left\{ \frac{1}{2R} I_{l+\frac{1}{2}}(\beta R) - \beta I'_{l+\frac{1}{2}}(\beta R) \right\} \right]. \quad (6)$$

Substituting (5) in (3) the first part of (5) produces the integral over x

$$\int_0^\infty \left\{ R J_{l+\frac{1}{2}}(xR) J_{l+\frac{1}{2}}(xr) - r \frac{l+1}{2l+1} J_{l+3/2}(xR) J_{l+3/2}(xr) \right. \\ \left. - r \frac{l}{2l+1} J_{l-\frac{1}{2}}(xR) J_{l-\frac{1}{2}}(xr) \right\} \frac{x dx}{x^2 + \beta^2}. \quad (7)$$

If

$$g(xR) = \frac{l}{2l+1} J_{l-\frac{1}{2}}(xR) - \frac{l+1}{2l+1} J_{l+3/2}(xR) = J'_{l+\frac{1}{2}}(xR) - \frac{1}{2xR} J_{l+\frac{1}{2}}(xR), \quad (8)$$

then from the recurrence relations it is easy to show that

$$-\frac{d}{dx} \{ xg(xR) J_{l+\frac{1}{2}}(xr) \} = x \left\{ R J_{l+\frac{1}{2}}(xR) J_{l+\frac{1}{2}}(xr) - r \frac{l+1}{2l+1} J_{l+3/2}(xR) J_{l+3/2}(xr) \right. \\ \left. - r \frac{l}{2l+1} J_{l-\frac{1}{2}}(xR) J_{l-\frac{1}{2}}(xr) \right\},$$

Hence integrating (7) by parts we obtain for it

$$-\int_0^\infty xg(xR) J_{l+\frac{1}{2}}(xr) \frac{2x}{(x^2 + \beta^2)^2} dx.$$

The integration over r can now be carried out, giving for the first part of (5) in the first part of (3) (Watson 1944 c)

$$-\frac{\pi\beta}{\sqrt{(\kappa R)}} \int_0^\infty \int_0^\infty \int_{(\beta\infty)} \frac{2}{\pi} \exp(-\alpha r) r J_{l+\frac{1}{2}}(\kappa r) J_{l+\frac{1}{2}}(xr) g(xR) \frac{2x^2}{(x^2 + \beta^2)^2} dr dx d\beta \\ = \frac{\pi\beta}{\sqrt{(\kappa R)}} \int_0^\infty \int_{(\beta\infty)} \frac{2\alpha}{\pi^2} \frac{1}{(\kappa x)^{3/2}} g(xR) \frac{dQ_l(X)}{dX} \frac{2x^2}{(x^2 + \beta^2)^2} dx d\beta \\ = \frac{2\alpha\beta}{\pi\kappa^2\sqrt{(R)}} \int_0^\infty \frac{1}{\sqrt{(x)}} g(xR) \frac{dQ_l}{dX} \left(\frac{\pi}{2x^2} + \frac{\tan^{-1} \beta/x}{x^2} + \frac{\beta}{x(x^2 + \beta^2)} \right) dx. \quad (9)$$

From (8) it is immediate that the first part of (6) in the second part of (3) will produce on integration over r

$$\frac{2\alpha\beta}{\pi\kappa^2\sqrt{(R)}} \int_0^\infty \frac{1}{\sqrt{(x)}} g(xR) \frac{dQ_l}{dX} \frac{x}{\beta(x^2 + \beta^2)} dx. \quad (10)$$

Adding (9) and (10) gives the first integral in the expression for M_{0n}^0 quoted in §2.

The second part of (5) when substituted in (3) will allow a similar partial integration in β and then the same integration over r giving

$$(i)^l \frac{2\alpha}{\kappa^2 \sqrt{(\beta R)}} g^*(\beta R) \frac{dQ_l}{dX_1} + (i)^l \frac{2\alpha\beta}{\kappa^2 \sqrt{(R)}} \int_0^\beta \frac{1}{\sqrt{(\beta)}} g^*(\beta R) \frac{dQ_l}{dX_1} \frac{1}{\beta^2} d\beta, \quad \dots (11)$$

where

$$X_1 = \frac{\alpha^2 + \kappa^2 - \beta^2}{2i\kappa\beta} \quad \text{and} \quad g^*(xR) = \frac{l}{2l+1} I_{l-\frac{1}{2}}(\beta R) + \frac{l+1}{2l+1} I_{l+\frac{3}{2}}(\beta R).$$

The second part of (6) when substituted in (3) integrates immediately over r and produces exactly the first term in (11) with opposite sign. Hence the second integral in the expression for M_{0n}^0 .

M_{0n}^1 is obtained in a similar manner from (1) by differentiating with respect to θ .

Asymptotic formulæ for large R .

If the integral I be expanded in descending powers of R and bearing in mind that β diminishes very rapidly to zero with increasing R , we find the asymptotic formula for $l=0$

$$I_{R \rightarrow \infty} = \frac{2}{\sqrt{(\kappa)}} \frac{1}{R^2} \int_0^\infty \exp(-\alpha r) r^{3/2} J_{\frac{1}{2}}(\kappa r) dr = \frac{4\sqrt{(2)}}{\sqrt{(\pi)}} \frac{\alpha}{(\alpha^2 + \kappa^2)^2} \cdot \frac{1}{R^2}$$

in which we may put $\alpha=1$.

Using this value in M_{0n}^0 and $W_0=-1$ the calculation of a_n follows classical lines giving for large Δ

$$a_n = \frac{8\kappa}{(1+\kappa^2)^2} K_0 \left(\frac{1+\kappa^2}{2v} \Delta \right).$$

Hence the method adopted here gives the classical result in the limit $\Delta \rightarrow \infty$.

REFERENCES.

WATSON, 1944, *Theory of Bessel Functions* (Cambridge: University Press),
(a) 365; (b) 429; (c) 412.

LXXIII. *Thermal Stresses in Turbine Blades.*

By M. J. LIGHTHILL (Manchester) and F. J. BRADSHAW (Farnborough) *.

[Received April 27, 1949.]

SUMMARY.

A theory of thermal stress in turbine blades is developed, on the assumption that at each point of the blade planform the stresses are approximately those that would be set up in a free infinite slab of uniform thickness equal to the thickness of the blade at that point. To justify this assumption it is shown that the thermal stresses in a wedge, assuming plane strain and infinite heat transfer coefficient, differ from those based on the approximate theory by terms which tend to zero uniformly with the wedge angle. More general arguments are also given in § 6. Consequences of the theory are that in cooling the maximum stress occurs at all times near the position of maximum thickness, but that in heating the largest stresses are initially near an edge, though as time goes on their position moves along the blade towards the position of maximum thickness, and their magnitude increases. Maximum stress is inversely proportional to thermal conductivity for the lower heat transfer rates, but is less sensitive to it at higher rates. Application to a specific material and comparison with experiment are given in § 7.

§1. INTRODUCTION.

WHEN a gas turbine is started up, or when the burners are switched off, the blades are suddenly surrounded by gas at a temperature considerably greater, or less, than their own. The resulting thermal stress frequently causes fracture in blades made of certain non-ductile materials (which might otherwise be suitable, owing to their strength at high temperatures); while repeated thermal shock may be the cause of the fatigue cracks in metal blades. For the brittle materials with which we are most concerned fracture is thought to occur when the maximum principal tensile stress exceeds a certain limit, called the breaking stress. Known solutions of the thermo-elastic equations can be applied to estimate this maximum stress, and one of these, the solution for an infinite slab, is given in §3. However, all these solutions apply to smooth bodies, and actually, at any rate in the heating process, the cracks often develop at the trailing edge, which is sharpened to an angle of as little as 6° in most blades. To assess the influence of an edge on the maximum thermal stress, we next investigate the stresses in an infinite wedge suddenly immersed in gas at a different temperature, assuming plane strain: these should simulate those near the trailing edge of an actual blade.

* Communicated by the Authors.

The result will depend on H , the heat transfer rate per unit area, per degree temperature difference between the solid surface and the gas outside the "temperature boundary layer." A simplifying approximation, though hardly adequate in the present problem, is to assume this temperature difference always zero: this is equivalent to taking $H = \infty$. The stresses may be expected to be greater in this case. In §4 the exact stress distribution for the wedge is given mathematical expression for the case $H = \infty$; this shows that the presence of the edge introduces no infinity into the stress field; and in §5 this expression is approximated, on the assumption that the wedge angle is small, by an expression simple to calculate. In §6, using the resulting knowledge of the relative orders of magnitude of quantities for small wedge angles, an approximate solution for general H , and for more general shapes of blade, is obtained. The numerical conclusions are discussed in §7.

§2. THE THERMO-ELASTIC EQUATIONS IN PLANE STRAIN.

Let a long homogenous cylinder, of thermal conductivity k , thermometric conductivity (diffusivity) κ , coefficient of linear expansion α , Young's modulus E , and Poisson's ratio σ , which is unstressed at temperature T_0 , be suddenly immersed, at time $t=0$, in gas at temperature $T=0$ (in an appropriate scale). Then in the absence of body forces and surface tractions along the generators, and on the assumption of certain applied temperature and stress distributions at the ends of the cylinder, the thermo-elastic equations may be solved on the assumption of "plane strain", which is that the strain components e_{xz} , e_{yz} and the stress components \widehat{xz} , \widehat{yz} vanish, while e_{zz} is a constant and the other stress and strain components and the temperature are functions of x , y only (the z -axis being that of the cylinder).

For on these hypotheses all the equations of equilibrium and of compatibility are satisfied identically except

$$\frac{\partial \widehat{xx}}{\partial x} + \frac{\partial \widehat{xy}}{\partial y} = 0, \quad \frac{\partial \widehat{xy}}{\partial x} + \frac{\partial \widehat{yy}}{\partial y} = 0, \quad \frac{\partial^2 e_{xx}}{\partial y^2} + \frac{\partial^2 e_{yy}}{\partial x^2} = \frac{\partial^2 e_{xy}}{\partial x \partial y}, \quad \dots \quad (1)$$

and all the stress-strain relations except

$$\left. \begin{aligned} E(e_{xx} - \alpha(T - T_0)) &= \widehat{xx} - \sigma(\widehat{yy} + \widehat{zz}), & E(e_{yy} - \alpha(T - T_0)) &= \widehat{yy} - \sigma(\widehat{zz} + \widehat{xx}), \\ E(e_{zz} - \alpha(T - T_0)) &= \widehat{zz} - \sigma(\widehat{xx} + \widehat{yy}), & Ee_{xy} &= 2(1 + \sigma)\widehat{xy}. \end{aligned} \right\} \quad \dots \quad (2)$$

Equations (1) and (2) are all satisfied if the stresses are represented in terms of an Airy stress function $\chi(x, y)$ by

$$\left. \begin{aligned} \widehat{xx} &= \partial^2 \chi / \partial y^2, & \widehat{xy} &= -\partial^2 \chi / \partial x \partial y, & \widehat{yy} &= \partial^2 \chi / \partial x^2, \\ \widehat{zz} &= \sigma \nabla^2 \chi + E(e_{zz} - \alpha(T - T_0)), \end{aligned} \right\} \quad \dots \quad (3)$$

where χ satisfies

$$\nabla^4 \chi + \frac{E\alpha}{1 - \sigma} \nabla^2 T = 0, \quad \dots \quad (4)$$

which must be solved on the assumption of zero surface tractions at the boundaries parallel to the z -axis. The temperature satisfies

$$\kappa \nabla^2 T = \partial T / \partial t, \quad (5)$$

with the boundary condition

$$HT + k \partial T / \partial n = 0 \quad (6)$$

expressing that the heat transfer rate from the gas to the solid is HT per unit area. There is a unique solution for these equations and boundary conditions. We may put $\chi = \chi_0 + \chi_1$, where χ_0 is a particular integral of (4)—thus

$$\chi_0 = - \frac{E\alpha\kappa}{1-\sigma} \int T \, dt \quad (7)$$

is one such, in view of (5)—and χ_1 is a biharmonic added to it to satisfy the boundary conditions.

The solution thus obtained would be exact if the ends of the cylinder were maintained at the temperature and normal stress distributions T and \widehat{zz} , with zero tangential stresses \widehat{xz} and \widehat{yz} : this would be achieved by thermal insulation and the maintenance of a constant distance between the two ends. In the real problem the temperature will reduce to zero more rapidly near the free end than on the two-dimensional theory and the stresses will presumably be smaller in this neighbourhood. It is harder to guess what will happen at the root, but conditions here may be less serious owing to a lower heat transfer rate. In any case plane strain will not be assumed in the final theory of §6.

§3. THE INFINITE SLAB.

If the body is an infinite slab, bounded by the planes $y = \pm a$, both T and χ will be independent of x and even functions of y . We may take as a particular integral of (4)

$$\chi_0 = - \frac{E\alpha}{1-\sigma} \int_0^y dy_1 \int_0^{y_1} T \, dy_2, \quad (8)$$

and add to it the even biharmonic

$$\chi_1 = - (\chi_0)_{y=a} - \frac{y^2 - a^2}{2a} \left(\frac{\partial \chi_0}{\partial y} \right)_{y=a} \quad (9)$$

to make the sum satisfy the conditions $\widehat{xy} = \widehat{yz} = 0$, or effectively $\chi = \partial \chi / \partial y = 0$, on $y = \pm a$. The non-vanishing stresses are then

$$\widehat{xx} = \frac{\partial^2 (\chi_0 + \chi_1)}{\partial y^2} = - \frac{E\alpha}{1-\sigma} T + \frac{E\alpha}{1-\sigma} \frac{1}{a} \int_0^a T \, dy, \quad . . . (10)$$

and $\widehat{zz} = \sigma \widehat{xx} + E(e_{zz} - \alpha(T - T_0))$. The most natural value of the constant

e_{zz} in this problem is that which makes $\widehat{xx}=\widehat{zz}$, so that there is no essential difference between the x - and z -directions: this requires that

$$e_{zz} = \frac{\alpha}{1-\sigma} \frac{1}{a} \int_0^a T dy - \alpha T_0 = e_{xx} \quad . \quad . \quad . \quad (11)$$

at each instant. In this case $\int_{-a}^a \widehat{xx} dy$ and $\int_{-a}^a \widehat{zz} dy$ are both zero and so by

Saint Venant's principle the stresses are those which would occur in a large finite slab with its perimeter free of surface traction. Other values of e_{zz} would require a uniform thrust in the z -direction to maintain them.

In cooling $T > 0$ throughout the field, and the maximum of the principal tensile stress (10) at each instant is attained at the boundary where T is least. In heating $T < 0$ and T is a minimum in the central section. Thus the maximum stresses in cooling and heating respectively are

$$\frac{E\alpha}{1-\sigma} \left(\frac{1}{a} \int_0^a T dy - (T)_{y=a} \right), \quad \frac{E\alpha}{1-\sigma} \left(\frac{1}{a} \int_0^a T dy - (T)_{y=0} \right). \quad . \quad (12)$$

The temperature distribution is best found by use of Heaviside's p -operator (which stands for the inverse of integration from 0 to t). The solution of

$$\kappa \partial^2 T / \partial y^2 = p(T - T_0), \quad . \quad . \quad . \quad (13)$$

if the boundary condition (6) holds at $y = \pm a$, is

$$T = T_0 \left(1 - \frac{\cosh qy}{\cosh qa + (qk/H) \sinh qa} \right), \quad . \quad . \quad . \quad (14)$$

where $q = \sqrt{(p/\kappa)}$. This may be interpreted by the expansion theorem in terms of the zeros of the denominator as a rapidly convergent series, except for small $\kappa t/a^2$, when the behaviour of (14) for large q should be used to obtain a quickly calculated expression. Omitting these, we obtain solely the maximum stresses. Putting $\beta = E\alpha T_0/(1-\sigma)$, a unit of stress which is positive in cooling and negative in heating, and $h = aH/k$, a non-dimensional quantity, the maximum stresses (12) become

$$\beta \frac{\cosh qa - (1/qa) \sinh qa}{\cosh qa + (qa/h) \sinh qa}, \quad \beta \frac{1 - (1/qa) \sinh qa}{\cosh qa + (qa/h) \sinh qa}. \quad . \quad (15)$$

Expanding in terms of the zeros of the denominator, these become

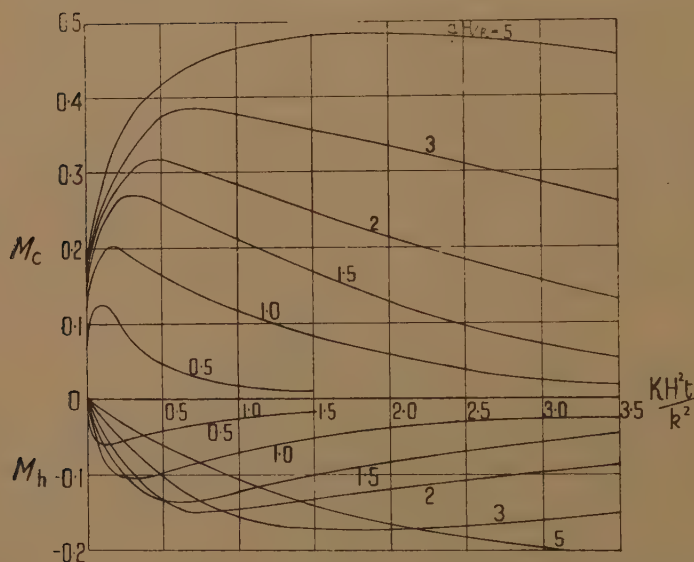
$$\beta \sum_{\lambda \tan \lambda = h} \frac{2 \sin \lambda (\sin \lambda - \lambda \cos \lambda)}{\lambda (\lambda + \sin \lambda \cos \lambda)} \exp(-\lambda^2 \kappa t/a^2),$$

$$\beta \sum_{\lambda \tan \lambda = h} \frac{2 \sin \lambda (\sin \lambda - \lambda)}{\lambda (\lambda + \sin \lambda \cos \lambda)} \exp(-\lambda^2 \kappa t/a^2). \quad . \quad . \quad (16)$$

On the other hand, for small $\kappa t/a^2$ they may be approximated by

$$\left. \begin{aligned} \beta \frac{1-(1/qa)}{1+(qa/h)} &= \beta - \beta \exp(h^2 \kappa t/a^2) \operatorname{erfc}(h\sqrt{(\kappa t)}/a) + \frac{\beta}{h} \left[1 - 2 \left(\frac{h^2 \kappa t}{\pi a^2} \right)^{\frac{1}{2}} \right. \\ &\quad \left. + \exp(h^2 \kappa t/a^2) \operatorname{erfc}(h\sqrt{(\kappa t)}/a) \right] = \beta \left[\left(\frac{2h}{a} \right) \left(\frac{\kappa t}{\pi} \right)^{\frac{1}{2}} + \dots \right], \\ \beta \frac{(-1/qa)}{1+qa/h} &= \frac{\beta}{h} \left[1 - 2 \left(\frac{h^2 \kappa t}{\pi a^2} \right)^{\frac{1}{2}} \right. \\ &\quad \left. + \exp(h^2 \kappa t/a^2) \operatorname{erfc}(h\sqrt{(\kappa t)}/a) \right] = -\beta \left[\frac{h \kappa t}{a^2} + \dots \right], \end{aligned} \right\} \dots (17)$$

Fig. 1.



Variation with time of centre and surface stresses of an infinite slab, plotted for different values of slab thickness.

the errors being $O(\exp(-a^2/\kappa t))$, $O(\exp(-a^2/4\kappa t))$ respectively: here $\operatorname{erfc} x$ is $(2/\sqrt{\pi}) \int_x^\infty \exp(-x^2) dx$.

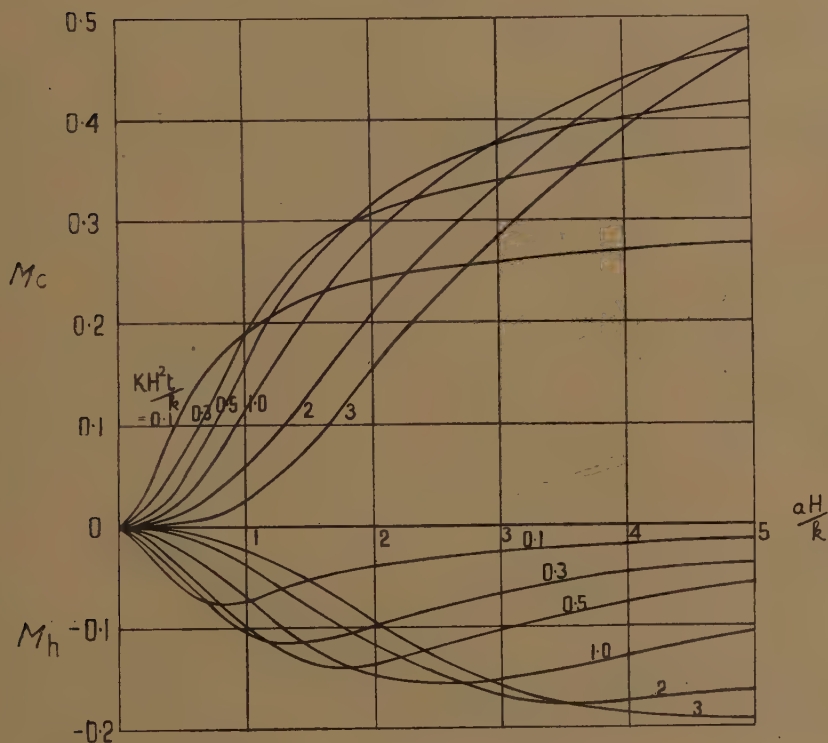
It is convenient to tabulate these functions in terms of $h=aH/k$ and of $h^2 \kappa t/a^2 = \kappa t H^2/k^2$, thus separating the influence of thickness and of time. The maximum stresses are thus written $\beta M_c(aH/k, \kappa t H^2/k^2)$ and $\beta M_h(aH/k, \kappa t H^2/k^2)$ in cooling and heating respectively. They are tabulated as functions of $\kappa t H^2/k^2$ for different aH/k in fig. 1 and as functions of aH/k for different $\kappa t H^2/k^2$ in fig. 2.

§4. THE INFINITE WEDGE ($H=\infty$).

For the wedge $0 < r < \infty$, $-\pi/2\nu < \theta < \pi/2\nu$, in cylindrical polars (r, θ, z), when $H=\infty$, the thermo-elastic problem of §2 contains no fundamental length- or time-scale. Hence the temperature and stresses can depend only on the non-dimensional variable $\rho=r/2\sqrt{(\kappa t)}$ and on θ . If $T/T_0=F(\rho, \theta)$, F is non-dimensional also and satisfies

$$\frac{\partial^2 F}{\partial \rho^2} + \left(2\rho + \frac{1}{\rho}\right) \frac{\partial F}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2 F}{\partial \theta^2} = 0, \quad \dots \quad (18)$$

Fig. 2.



Variation with slab thickness of centre and surface stresses of an infinite slab, plotted at different times.

where $F=0$ on $\theta=\pm\pi/2\nu$, while $F\rightarrow 1$ as $\rho\rightarrow\infty$. By (7) we may put

$$\chi_0 = -\frac{E\alpha T_0}{1-\sigma} \int_0^e F(\rho, \theta) \frac{d(\kappa t)}{d\rho} d\rho = 4\beta\kappa t \left(\frac{1}{2}\rho^2 \int_0^e \rho^{-3} F(\rho, \theta) d\rho \right) = 4\beta\kappa t \phi_0(\rho, \theta). \quad \dots \quad (19)$$

If $\chi=\chi_0+\chi_1=4\beta\kappa t(\phi_0(\rho, \theta)+\phi_1(\rho, \theta))=4\beta\kappa t\phi$, the stresses are

$$\begin{aligned} \widehat{rr} &= \beta \left(\frac{1}{\rho^2} \frac{\partial^2 \phi}{\partial \theta^2} + \frac{1}{\rho} \frac{\partial \phi}{\partial \rho} \right), & \widehat{r\theta} &= -\beta \frac{\partial}{\partial \rho} \left(\frac{1}{\rho} \frac{\partial \phi}{\partial \theta} \right), & \widehat{\theta\theta} &= \beta \frac{\partial^2 \phi}{\partial \rho^2}, \\ \widehat{zz} &= \sigma\beta\nabla^2 \phi + E(e_{zz} - \alpha(T-T_0)). \quad \dots \quad (20) \end{aligned}$$

The boundary conditions on ϕ are that $\widehat{r\theta}=\widehat{\theta\theta}=0$ on $\theta=\pm\pi/2\nu$; and if $\phi=O(\rho^2)$ as $\rho\rightarrow 0$, these imply $\phi=\partial\phi/\partial\theta=0$.

A series of cosines vanishing at $\theta=\pm\pi/2\nu$ is appropriate for the even function F . Since

$$1=\frac{4}{\pi}\sum_{n=0}^{\infty}(-1)^n\frac{\cos(2n+1)\nu\theta}{2n+1}\left(-\frac{\pi}{2\nu}<\theta<\frac{\pi}{2\nu}\right), \quad \dots (21)$$

the series for F (which $\rightarrow 1$ as $\rho\rightarrow\infty$) is

$$F=\frac{4}{\pi}\sum_{n=0}^{\infty}(-1)^n\frac{\cos(2n+1)\nu\theta}{2n+1}F_{(2n+1)\nu}(\rho), \quad \dots (22)$$

where $F_m(\rho)$ is that solution of

$$F_m''(\rho)+(2\rho+\rho^{-1})F_m'(\rho)-m^2\rho^{-2}F_m(\rho)=0, \quad \dots (23)$$

which is bounded at $\rho=0$ and tends to unity as $\rho\rightarrow\infty$. The solutions of (23) are the confluent hypergeometric functions

$$\rho^m{}_1F_1(\tfrac{1}{2}m; m+1; -\rho^2), \quad \rho^{-m}{}_1F_1(-\tfrac{1}{2}m; -m+1; -\rho^2). \quad \dots (24)$$

The latter is not bounded at $\rho=0$. The multiple of the former which tends to unity as $\rho\rightarrow\infty$ is

$$F_m(\rho)=\frac{(\frac{1}{2}m)!}{m!}\rho^m{}_1F_1(\tfrac{1}{2}m; m+1; -\rho^2)\sim 1-\frac{\frac{1}{4}m^2}{1!\rho^2}+\frac{\frac{1}{4}m^2(\frac{1}{4}m^2-1^2)}{2!\rho^4}-\dots \quad \dots (25)$$

Hence, by (19),

$$\phi_0(\rho, \theta)=\frac{4\nu}{\pi}\sum_{n=0}^{\infty}(-1)^n\Phi_{(2n+1)\nu}(\rho)\cos(2n+1)\nu\theta, \quad \dots (26)$$

where

$$\Phi_m(\rho)=\frac{\rho^2}{2m}\int_0^{\rho}\rho^{-3}F_m(\rho)d\rho=\frac{(\frac{1}{2}m-2)!}{m!}\rho^m{}_1F_1(\tfrac{1}{2}m-1; m+1; -\rho^2). \quad \dots (27)$$

By (26), ϕ_0 vanishes on $\theta=\pm\pi/2\nu$, but $\partial\phi_0/\partial\theta$ does not. Hence the biharmonic $\phi_1(\rho, \theta)$ must satisfy $\phi_1=0$ and

$$\frac{\partial\phi_1}{\partial\theta}=\pm\frac{4\nu}{\pi}\sum_{n=0}^{\infty}(2n+1)\nu\Phi_{(2n+1)\nu}(\rho)=\pm f(\rho) \quad \dots (28)$$

on $\theta=\pm\pi/2\nu$.

Now $f(\rho)$ is a double series of multiples of positive powers of ρ . It remains uniformly convergent, as do its derivatives, in any bounded range of ρ , even when the coefficient of each power is replaced by its modulus, since for example

$${}_1F_1(\tfrac{1}{2}m-1; m+1; \rho^2)\rightarrow e^{\frac{1}{2}\rho^2} \text{ as } m\rightarrow\infty. \quad \dots (29)$$

Hence the ϕ_1 obtained from $f(\rho)$ by replacing each power ρ^N therein by

$$\rho^N\frac{\cos(N-2)\theta\cos N\pi/2\nu-\cos N\theta\cos(N-2)\pi/2\nu}{(N-1)\sin\pi/\nu+\sin(N-1)\pi/\nu}, \quad \dots (30)$$

is a biharmonic which satisfies the required boundary conditions. It introduces no infinite stresses into the field.

We proceed at once to the approximation for small π/ν but may remark that if an estimate of the stress distribution for larger wedge angles were desired, the term $\sin(N-1)\pi/\nu$ in the denominator of (30) might as a rough approximation be omitted, in which case ϕ_1 would have an expression in terms of the indefinite integral of $\rho^{-2}f(\rho)$.

§5. APPROXIMATION TO THE SOLUTION OF §4 FOR SMALL WEDGE ANGLE.

Since in the functions $\Phi_m(\rho)$ occurring in (26) m takes the values $\nu, 3\nu, 5\nu, \dots$, and since in the application described in §1 ν is about 30, it is reasonable to replace $\Phi_m(\rho)$ by an asymptotic form for large m , provided that the latter is valid uniformly for large ρ (for it may happen that the values of ρ for which the stresses are largest become large as ν does).

Jeffreys's method is used. In expression (27) for $\Phi_m(\rho)$ put $\rho^2 = mx$. The resulting function of x satisfies a differential equation with m as a parameter, and a hypothetical asymptotic form

$$\Phi_m(\rho) \sim \frac{1}{m^2} \exp(m\omega(x)) \left[y_0(x) + \frac{1}{m} y_1(x) + \dots \right] \quad (31)$$

may be substituted in this differential equation. First order equations for $\omega(x)$, $y_0(x)$, $y_1(x)$, etc., result from equating powers of m . But by the behaviour of $\Phi_m(\rho)$ for small ρ , $\omega(x) \sim \frac{1}{2}(1 + \log \frac{1}{2}x)$ and $y_0(x) \rightarrow 2^{-3/2}$ as $x \rightarrow 0$. With these boundary conditions the equations for ω and y_0 have the unique solutions

$$\left. \begin{aligned} \omega(x) &= \frac{1}{2} \left[\sqrt{1+x^2} - x + \log \frac{x}{1+\sqrt{1+x^2}} \right], \\ y_0(x) &= \left(\frac{x + \sqrt{1+x^2}}{2} \right)^{3/2} (1+x^2)^{-1/4}. \end{aligned} \right\} \quad (32)$$

The crucial point is that $\omega(x)$ increases from $-\infty$ at $x=0$ monotonically to zero at $x=\infty$. When $\omega(x)$ is small in magnitude x is large, $\omega(x) \simeq -1/4x$ and $y_0(x) \simeq x$. Now, by (26) and (31), ignoring a factor $1+O(1/\nu)$,

$$\begin{aligned} \phi_0(\rho, \theta) &= \frac{4\nu}{\pi} \sum_{n=0}^{\infty} (-1)^n \frac{\cos(2n+1)\nu\theta}{(2n+1)^2\nu^2} \\ &\times \exp \left[(2n+1)\nu\omega \left(\frac{\rho^2}{(2n+1)\nu} \right) \right] y_0 \left(\frac{\rho^2}{(2n+1)\nu} \right). \quad (33) \end{aligned}$$

The numbers $(2n+1)\nu$ take values about 30, 90, 150, \dots . It is clear that the only values of n which contribute to ϕ_0 or its derivatives, terms of a magnitude worth considering are those for which $\omega(x)$ at $x = \rho^2/(2n+1)\nu$ is small in magnitude. But for these $\omega(x)$ can be approximated by $-1/4x$ and $y_0(x)$ by x . Since replacing ω and y_0 by these simple forms will also

leave small the terms which were already small in (33), the approximation which it affords, namely

$$\phi_0(\rho, \theta) = \frac{4}{\pi} \left(\frac{\rho}{\nu} \right)^2 \sum_{n=0}^{\infty} (-1)^n \frac{\cos (2n+1)\nu\theta}{(2n+1)^3} \exp(-(2n+1)^2\nu^2/4\rho^2) \quad (34)$$

is valid for large ν . The error can be estimated by splitting up the difference between (33) and (34) into terms with $(2n+1)\nu \leq 2\rho\sqrt{(\log \rho)}$, for which $\omega(x) + 1/4x = O(1/x^2)$, and those with $(2n+1)\nu > 2\rho\sqrt{(\log \rho)}$, for which the inequality $\omega(x) \leq \frac{1}{2} \log [x/(x+\frac{1}{2})]$ may be used: the error is found to be $O(\nu^{-1} \log^2 \nu)$ as $\nu \rightarrow \infty$.

The ϕ_0 of (34) is a function of ρ/ν and $\nu\theta$ only, and is otherwise independent of ν . A ϕ_1 can now be obtained with $\phi_1=0$ and $\partial(\phi_0+\phi_1)/\partial\theta=0$ on $\theta=\pi/\nu$, so that ϕ_1 is also a function of ρ/ν and $\nu\theta$ only (plus terms which are smaller, for large ν). But for such functions $\partial^2\phi/\partial\rho^2$ and $\rho^{-1}\partial\phi/\partial\rho$ are smaller by a factor of order ν^{-2} compared with derivatives like $\rho^{-2}\partial^2\phi/\partial\theta^2$. Hence the equation satisfied by ϕ_1 becomes $\partial^4\phi_1/\partial\theta^4=0$ and the solution is

$$\phi_1 = -\frac{\nu}{\pi} \left(\theta^2 - \frac{\pi^2}{4\nu^2} \right) \left(\frac{\partial\phi_0}{\partial\theta} \right)_{\theta=\pi/2\nu} \dots \dots \dots (35)$$

Further, by (20), the stresses $\widehat{r\theta}$ and $\widehat{\theta\theta}$ are small compared with \widehat{rr} and \widehat{zz} , which are comparable, the former being given by

$$\begin{aligned} \widehat{rr} &= \frac{\beta}{\rho^2} \frac{\partial^2(\phi_0+\phi_1)}{\partial\theta^2} = \beta \left[\frac{8}{\pi^2} \sum_{n=0}^{\infty} \frac{\exp(-(2n+1)^2\nu^2/4\rho^2)}{(2n+1)^2} \right. \\ &\quad \left. - \frac{4}{\pi} \sum_{n=0}^{\infty} (-1)^n \cos (2n+1)\nu\theta \frac{\exp(-(2n+1)^2\nu^2/4\rho^2)}{2n+1} \right] \quad (36) \end{aligned}$$

§6. APPROXIMATE SOLUTION FOR GENERAL H AND FOR ANY THIN BLADE.

The conclusion of §5, is that, for an infinite wedge in plane strain with $H=\infty$, the stress distribution can be approximated for small wedge angle π/ν by a distribution depending on ρ/ν and $\nu\theta$ only, so that radial derivatives of quantities are small compared with transverse derivatives: further, no singularity appears at the edge.

It will now be assumed that a similar result holds for any blade and for any value of H ; namely that, when the blade is thin, the stress distribution can be approximated by one for which derivatives of quantities tangential to the middle surface are small compared with derivatives normal to it. Then the thermo-elastic equations and boundary conditions at any point of the blade planform become those appropriate to the infinite slab solution of §3, with the thickness of the slab equated to the blade thickness at that point. And in fact the chordwise stress (36) agrees with that obtained from the solution of §3 by putting $H=\infty$, $y=r\theta$ and $a=r\pi/2\nu$ (since the roots of $\lambda \tan \lambda = aH/k$ then become $\lambda = (2n+1)\frac{1}{2}\pi$ where $n=0, 1, 2, \dots$).

The hypothesis of plane strain must now be dropped: it should be observed that for the wedge a thrust in the z -direction, whose resultant across the thickness of the blade varies with x , is required to maintain it. In fact, by (20) $\widehat{zz} = \sigma \widehat{rr} + E(e_{zz} - \alpha(T - T_0))$; hence, since the resultant of \widehat{rr} is zero, that of \widehat{zz} is

$$- \int_{-\pi/2\psi}^{\pi/2\psi} E\alpha Tr d\theta + \text{constant}, \dots \dots \dots (37)$$

which is not independent of r .

It is reasonable to suppose that for a free blade, as was shown for the free slab in §3, $\widehat{xx} = \widehat{zz}$, so that both have zero resultant over the thickness of the blade. Then the approximate equilibrium of any large area of blade follows from a case of Saint Venant's principle, which states that the effect of unresolved stresses, statically equivalent to zero, acting at a thin strip, falls off exponentially as the distance from the strip of a point in a thin plate bounded by it becomes large compared with the thickness: further, that of any small area of blade follows from the slow variation of the blade thickness and so of \widehat{xx} and \widehat{zz} . Slight errors will be introduced by the failure of this requirement near the rounded leading edge.

The validity of the theory can be estimated for a wedge of finite chord c and large aspect ratio A . The stresses will be those of §5 if tractions \widehat{rr} , \widehat{zz} are maintained on $r=c$, $z=\pm\frac{1}{2}Ac$ respectively. Except at points whose distance from these boundaries is comparable with the thickness the effect of these tractions may be replaced by their resultants across the thickness of the blade, namely zero on $r=c$ and (37) on $z=\pm\frac{1}{2}Ac$. Now the theory of long cylinders of arbitrary cross-section, bent by forces along the generators applied at the two ends, gives that except near the ends the only stress is a tension or thrust along the generators, whose variation normal to the generators is linear and statically equivalent to the applied stresses. So in this problem, when the tractions on $z=\pm\frac{1}{2}Ac$ are relaxed, the variation of \widehat{zz} from \widehat{xx} (namely (37)) is reduced everywhere by subtracting a linear function equimomental with it. (The further removal of thermal insulation from the boundaries will also not alter the solution at some distance from them.) Probably for blades with all their edges tapered and not of large aspect ratio the approximation will be better still.

The maximum stress reached at time t at a point where the thickness is τ is $\beta M_c(\tau H/k, \kappa t H^2/k^2)$ in cooling (attained at the surface) and $\beta M_h(\tau H/k, \kappa t H^2/k^2)$ in heating (attained at the centre.) The absolute maximum of either will be the maximum of the function for all $\kappa t H^2/k^2$ and for $\tau H/k \leq \tau_{\max} H/k$. It is attained for $\tau = \tau_{\max}$, i.e. at the position of maximum thickness. But in cases where this maximum stress exceeds the breaking stress, fig. 2 shows that in heating, but not in cooling, the breaking stress will be first reached (in point of time) for a value of $\tau \leq \tau_{\max}$. For small $\kappa t H^2/k^2$ the position of the largest stress is near an edge, and it moves towards the thicker parts of the blade as t increases: at the same time its value increases.

§7. NUMERICAL CONCLUSIONS.

The principal remaining source of error in the above theory is the assumed constancy of H . This could be partially removed by allowing the value of H to vary across the blade planform (thus its leading edge value may exceed its value over rear portions by a factor of 5) and assuming that the resulting changes are only local; but it is certain that H also varies with time. Similarly, ignorance of even a rough value of H is the principal obstacle to numerical application of the theory. Other difficulties are that the breaking stress may be different when stress changes occur rapidly than in the simple static tests used to determine it, and that thermal conductivity varies with temperature and is at best not precisely known.

Mr. W. Watt of the Royal Aircraft Establishment has subjected blade shapes of sintered alumina separately to hot and cold shock. They were then visually examined for cracks (it is open to criticism whether this is a suitable test for failure). Blades at room temperature, suddenly inserted in a hot tunnel, appeared to crack if the tunnel temperature exceeded 700°C . Blades at temperature exceeding 400°C . appeared to crack when inserted in an air jet at room temperature.

Using c.g.s. units the blade material satisfied approximately $\tau_{\max}=0.5$, $k=0.004$, $\alpha=7.4.10^{-6}$, $E=3.7.10^{12}$, $\sigma=0.3$, breaking stress $=2.4.10^9$. Very rough estimates of H , taking into account the tunnel speeds which were different in heating and cooling, were 0.005 and 0.008 respectively, but these were based on empirical formulæ used far outside their range of applicability. (An experiment on an alumina rod to determine H in heating gave 0.01 though again this would only be an average value.) The estimates of H give $aH/k=0.31$, 0.51 in heating and cooling, and so the largest numerical values of the non-dimensional stresses M_h and M_c are read from fig. 2 as 0.042, 0.127. Multiplying their reciprocals by $[(1-\sigma)/E\alpha]$ times the breaking stress, we find that the temperature changes for which the theory predicts failure are 1460°C . in heating, 480°C . in cooling. The latter is in reasonable agreement with experiment, the former in excess. But a larger value of H such as the subsidiary experiment suggested would reduce the temperature change, roughly in inverse proportion, so that the margin of error is not greater than might be expected. Further experimental information on heat transfer and thermal shock is needed to make more exact conclusions possible. Meanwhile we believe that our theory gives a useful guide to the manner in which thermal failure occurs.

ACKNOWLEDGEMENT

This paper is published with the permission of the Chief Scientist, Ministry of Supply. The figures are copyright reserved and reproduction is with the permission of the Controller of H.M. Stationery Office.

LXXIV. *Notices of New Books and Periodicals received.*

Viscometry. By Dr. A. C. MERRINGTON. [Pp. 142.] (London: Arnold, 1949.) Price 16s.

THE author, in his preface, states that he has attempted to produce a concise and up-to-date account of most of the methods available for measuring viscosity with some discussion of their advantages and limitations. He has succeeded admirably and has added for good measure a copious bibliography and some useful numerical tables. N. T.

Radio-Frequency Heating. By L. HARTSHORN. (Allen & Unwin.) Price 21s.

THIS book deals with both induction and dielectric heating in a very practical manner, with emphasis on industrial rather than laboratory applications. The theory given is fairly elementary, but presented in a way giving a very satisfactory and straightforward picture of the processes involved, and is in any case probably sufficiently accurate for most practical purposes. There are many illustrations and a good deal of numerical information on the properties of substances likely to be encountered, and on the powers, frequencies, etc. required in typical applications. The bibliography is quite extensive.

D. F. G.

Cardinal Algebras. By ALFRED TARSKI. (Oxford University Press.) Price 50s.

STUDY of the arithmetic of cardinal numbers has led the author to the consideration of new kinds of abstract algebras. These have two fundamental operations, one of binary addition and one of addition of the terms of an infinite sequence, which satisfy certain elementary postulates of closure, commutativity, associativity, etc. The first part of this book is devoted to a systematic study of the properties of these algebras; in the second part connections with well-known algebraic systems such as semigroups and lattices are discussed.

Pulses and Transients in Communication Networks. By COLIN CHERRY. (Chapman & Hall, Ltd.) Price 32s.

THE difficulty with this subject is that while many of the fundamentals are fairly easily grasped and visualised, a rigorous mathematical treatment tends to obscure the simple physical concepts and is rather a matter for the pure mathematician. Thus most of the books available are either unapproachable by the non-mathematician or else contain theoretical processes of a highly questionable character. The author of this book is to be congratulated in having achieved a good compromise between these extremes, and thus having provided an introduction to the subject which is readable by anyone familiar with elementary steady-state A.C. theory. The analysis uses real Fourier integrals for the most part; Laplace transforms and contour integrals are not introduced. As stated in its sub-title, it is intended for television and radar engineers, and there is a liberal use of diagrams of vectors, waveforms and frequency spectra, the emphasis throughout being on useful practical examples and visualisable concepts rather than mathematical niceties. A catholic and up-to-date bibliography is provided for those wishing to explore the subject further.

D. F. G.

Divergent Series. By G. H. HARDY. (Oxford University Press: Geoffrey Cumberlege.) 1949. Price 30s. net.

DIVERGENT Series were a mystery during the XVIIIth century, and a despised and largely discarded tool during the XIXth century. It was only in recent times that Cesàro, Hölder, Borel and Toeplitz started to turn the study of divergent series into a profitable scientific pursuit. Nobody has done more to explore this new field than Hardy himself, largely in collaboration with Littlewood.

The present book gives an up-to-date account of the theory. In spite of its encyclopaedic character it is eminently readable. Hardly any involved problems appear to be left, and the reviewer has not seen a single proof in the book which strikes him as unduly long and complicated.

After two interesting historical chapters, the book deals with all the usual methods of summation, and their connection by Abelian and Tauberian theorems, including Wiener's theory. There are also chapters on multiplication of series, Hausdorff means and on the Euler-Maclaurin formula; but no other asymptotic developments are considered.

Undoubtedly the book will be the textbook on divergent series for many years to come. Whether it will also attract applied mathematicians appears more doubtful.

The thanks of the mathematical world are due to Dr. Bosanquet and Dr. Egglestone who have seen the book through the press. H. H.

The Mathematics of Great Amateurs. By J. L. COOLIDGE. (Oxford University Press: Geoffrey Cumberlege.) 1949. Price 21s. net.

THE author gives an account of the work of the following sixteen amateurs: Plato, Omar Khayyám, Pietro dei Franceschi, Leonardo da Vinci, Dürer, Napier, Pascal, Arnauld, De Witt, Hudde, Brouncker, L'Hospital, Buffon, Diderot, Horner, Bolzano. The author felt it was wrong to call men like Archimedes or Fermat amateurs, and so selected his teams which, excepting Pascal, contains no mathematician of the front rank.

History of mathematics is never easy to read, but when the task is to study second rate achievements, it becomes grim indeed. In spite of the author's scholarly treatment and lucid exposition of his subject, the book will probably appeal only to a limited number of readers. H. H.

Anuario del Observatorio Astronomico de Madrid, 1949, Madrid, Spain.

THIS volume gives a detailed account of the different Calendars and also of the Astronomical Ephemerides for 1949; those observable in Spain are considered in detail. An interesting article on distances in Astronomy, by R. Carrasco, is published at the end of the volume.

L'Electromagnetisme, Hier et Aujourd'hui. By EDWARD BAUER, Professor at the Sorbonne. (Editions, Albin Michel, Paris.)

THIS interesting little book of essays traces the history of electromagnetism from the earliest times, through the epoch of Newton, to the present day. The subject is taken to include relativity and much of atomic theory, and the author treats particularly the influence of physical theories on the philosophical and ethical ideas of the periods treated.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

LXXV. *The Numerical Reduction of Non-Singular Matrix Pencils.*

By J. M. HAMMERSLEY,

Lectureship in the Design and Analysis of Scientific Experiment,
University of Oxford †.

[Received April 26, 1949.]

SUMMARY.

The problem of reducing a non-singular matrix pencil $\mathbf{L} \equiv l_1 \mathbf{L}_1 + l_2 \mathbf{L}_2$, $|\mathbf{L}| \neq 0$, to canonical form, and of finding the latent roots $l_1 : l_2$ and root-vectors \mathbf{x} that satisfy $\mathbf{L}\mathbf{x} = \mathbf{0}$, arises in many branches of applied mathematics. The most familiar of these applications is the analysis of small vibrations of a dynamical system about an equilibrium position by means of Lagrangian frequency equations; while other applications include the conditioned stationary values of multivariate functions, approximations to certain integral equations, orthogonal mean square regressions in multivariate statistics, linear operational equations with constant coefficients, the determination of matrix commutants, matrix interpolation, and the growth behaviour of animal populations. The stock of current reduction procedures embraces the determinantal method, the Duncan-Collar-Aitken-Hotelling method, the Morris escalator, Samuelson's method, Horst's method, and the Cayley-Hamilton method. For one reason or another, each of these methods proves somewhat unsatisfactory—either it demands considerable computing labour, or it is unsuitable for automatic computers, or it only solves certain sections of the problem, or it is restricted to special cases such as symmetric matrices or distinct latent roots. In this paper I propound a new method, which I consider an improvement on current procedures. This new method is an extension of the Cayley-Hamilton method, based upon the general theory of collineations. It is quite general, not requiring modification in special cases; it provides the complete solution of all the latent roots and all the root-vectors and/or invariant space bases; it is suitable for punched card analysis and automatic computers—with the exception of a simple set of divisions, ring operations alone suffice for the whole computation—and it calls for rather less computing labour than current methods.

§1. THE COLLINEATORY FORM.

If \mathbf{L} is non-singular, we may by a change of base $\{l_1, l_2\} \rightarrow \{m_1, m_2\}$ transform \mathbf{L} into $\mathbf{M} \equiv m_1 \mathbf{M}_1 + m_2 \mathbf{M}_2$ where neither \mathbf{M}_1 nor \mathbf{M}_2 is singular. However, to permit greater flexibility and to avoid any preliminary

† Communicated by the Author.

transformation wherever possible, we shall only suppose that not both M_1 and M_2 are singular. Employing non-homogeneous coordinates for the base $\{l_1, l_2\}$ we shall write $Lx=0$ as

$$(A-\lambda B)x=0, \quad (1)$$

where A and B are square matrices of the n th order not both singular. (In almost all cases $L_1=A$ and $L_2=B$; but should the original L_1 and L_2 have both been singular, (1) will have been obtained by a change of the base $\{l_1, l_2\}$, and then A, B will be linear combinations of L_1, L_2 .) We may suppose that B is non-singular: for, if A were non-singular and B were singular, the characteristic equation could not have a zero root λ ; and we should be at liberty to write $1/\lambda$ for λ and to interchange A and B in our notation. Thus B^{-1} exists; and premultiplying by B^{-1} , (1) reduces to the form

$$(T-\lambda I)x=0, \quad (2)$$

where

$$T=B^{-1}A. \quad (3)$$

We call (2) the collineatory form; for it is the equation for the self-corresponding points under a collineation T of the $(n-1)$ -dimensional projective space S_{n-1} . In general, T is a non-singular or a singular collineation; the only modification, necessary for convenience in the latter case, being to augment S_{n-1} by the null vector to which all vectors correspond.

§2. REVIEW OF CURRENT METHODS.

The Determinantal Method.

In this method we evaluate $\phi(\lambda)\equiv |A-\lambda B|$ for $(n+2)$ arbitrary equidistant values of λ , say $\lambda=\mu_1, \mu_2, . . . , \mu_{n+2}$, by any one of the usual methods (*e.g.* Aitken 1932, Hartley 1946, Laderman 1948). All these methods possess running checks; and there is no foundation for the statement (Morris and Head 1944) that the determinantal method cannot be checked until the insertion of the latent roots. Then, when ϕ has real roots, we difference ϕ employing the check $\Delta^{n+1}\phi=0$, subtabulate it as desired, and solve $\phi=0$ by inverse interpolation (Comrie 1936, 1937). Should the roots of ϕ be complex, we may use the Graeffe method of solution (Brodetsky and Smeal 1924, Whittaker and Robinson 1940) or some method of vector iteration for the two-dimensional vector $\lambda=\alpha+i\beta=\{\alpha, \beta\}$ (*e.g.* Hartree 1948). We finally solve $(A-\lambda_i B)x_i=0$ for x_i , where λ_i is a latent root (Doolittle 1878, Aitken 1932, Dwyer 1941, Hartley 1946, Laderman 1948). When x_i is unique (apart from a scalar multiple), it is the root-vector corresponding to the latent root λ_i ; while, if it contains further disposable unknowns, it is decomposable into the base of the invariant space corresponding to λ_i .

The determinantal method is perfectly general: it needs no modifications in special cases (*e.g.* multiple latent roots). The computing is, however, heavy because of the $2(n+1)$ determinantal reductions.

The Duncan-Collar Method.

This method (Duncan and Collar 1934) applies to the collineatory form (2) when the latent roots are real and all of distinct magnitudes. If λ_1 is the unique root of greatest magnitude, it may be shown that $\mathbf{T}^m \sim \lambda_1^m \mathbf{T}_\infty$ as $m \rightarrow \infty$, where \mathbf{T}_∞ is of rank 1 and has columns proportional to the root-vector \mathbf{x}_1 associated with λ_1 . In particular, if \mathbf{u} is an arbitrary vector, $\mathbf{T}^m \mathbf{u} \rightarrow t_m \mathbf{x}_1$ where $t_{m+1}/t_m \rightarrow \lambda_1$ as $m \rightarrow \infty$. The equation $\mathbf{T}_\infty \mathbf{x} = \mathbf{0}$ is equivalent to a single linear equation, which may be used to eliminate one variable from the original equations (2). This process, known as "deflating" the original system, yields a reduced system from which λ_1 and \mathbf{x}_1 are absent: so that a repetition of the above steps gives the latent root of second largest magnitude and its root-vector; and so on until all roots and root-vectors are determined. In practice some suitably high power \mathbf{T}^m has to serve as an approximation to \mathbf{T}_∞ ; and the adequacy of this approximation and the success of the method depends considerably on the ratio separation of the latent roots. Even when this separation is favourably large, the number of matrix multiplications and successive substitutions constitute a heavy computing burden, especially since a large number of digits have to be retained in the early stages to ensure accuracy in the final results.

There is, however, a special case for which the Duncan-Collar method is peculiarly suitable. In the application to vibrating dynamical systems, the root λ_1 of greatest magnitude leads to the fundamental oscillation. If interest centres merely upon this fundamental and its normal mode, and not upon any of the overtones, then the mere evaluation $\mathbf{T}^m \mathbf{u}$ as $m \rightarrow \infty$ provides all that is wanted. If the convergence of this sequence is disappointingly slow, a few preliminary squaring operations $\mathbf{T}^2, \mathbf{T}^4, \dots, \mathbf{T}^{2^M} = \mathbf{Z}$, say, followed by $\mathbf{Z}^m \mathbf{u}$ ($m=1, 2, \dots$) will accelerate the process. Similarly in the orthogonal mean square regression problem one is concerned with the root of smallest magnitude, and the same procedure suffices if we consider the inverted equation in $1/\lambda$. In any iteration process of this sort, one needs to know where to stop. This is not an easy question to answer: for the mere fact that t_{m+1}/t_m nearly equals t_m/t_{m-1} is no indication that either ratio approximates to λ_1 . Hotelling (1943 a, 1943 b) investigates this problem, and gives upper and lower bounds for the estimates of λ_1 .

Aitken (1937 b) also sets out an ingenious δ^2 -process for accelerating the convergence to λ_1 . Writing $\mathbf{w}_i = \mathbf{T}^i \mathbf{u}$, the sequence of vectors $\dots, \mathbf{w}_{i-1}, \mathbf{w}_i, \mathbf{w}_{i+1}, \dots$, yields the scalar functions

$$\begin{aligned} & \dots \dots \dots \\ f(2i-2) &= \mathbf{w}'_{i-1} \mathbf{w}_{i-1} \\ f(2i-1) &= \mathbf{w}'_{i-1} \mathbf{w}_i \\ f(2i) &= \mathbf{w}'_i \mathbf{w}_i \\ f(2i+1) &= \mathbf{w}'_i \mathbf{w}_{i+1} \\ f(2i+2) &= \mathbf{w}'_{i+1} \mathbf{w}_{i+1} \\ & \dots \dots \dots \end{aligned}$$

Let

$$\begin{aligned}\psi(m) &= f(m+1)/f(m), \\ \delta^2\psi(m) &= \psi(m+1) - 2\psi(m) + \psi(m-1), \\ \psi_2(m) &= \psi(m+1)\psi(m-1) - \psi(m)\psi(m).\end{aligned}$$

Then, as Aitken proves, the ratio $\psi_2(m)/\delta^2\psi(m)$ converges to λ_1 (as $m \rightarrow \infty$) much more rapidly than $\psi(m)$ does.

The methods described in this section also apply to the case when there are multiple latent roots and/or complex conjugate roots. Aitken (1937 b) gives the appropriate formulæ. In these more general cases, the asymptotic behaviour of $\mathbf{T}^m \mathbf{u}$ is no longer a simple geometric progression; and the task of recognizing when the asymptotic state has been virtually reached is scarcely delegable to routine computers.

The Morris Escalator.

In their account of the escalator method, Morris and Head (1944) wrote:—"It was decided to keep the analysis on simple algebraic lines. It may happen that those well versed in the matrix notation will find it easier to follow the process by working it out in that notation, but the authors set out to cater for those who, like the authors themselves, have minds which are not readily attuned to the matrix notation." Weatherburn (1917), in a plea for the use of vector notation, states the opposite viewpoint:—"It has often been remarked that nothing can be accomplished by vector methods that cannot also be done by Cartesian analysis If we had to deal only with minds of special mathematical ability and analytical insight, this conclusion might be accepted. But with the average student so much of his attention is occupied in dealing with the complex array of symbols . . . in Cartesian analysis, that he is unable to grasp the inner meaning of the work." A matrix proof of the escalator method is justified on these grounds alone; and is certainly recommended when in addition we see that, despite prolix algebraic expressions and a number of unworked manipulative stages, the original proof is confined to the particular case $\nu=3$ (in the notation adopted below). A statement of the method and a proof for general values of ν follows.

While the escalator method can be extended to the general case, I shall (following the initial statement of Morris and Head) treat only of the case where \mathbf{A} and \mathbf{B} in (1) are symmetrical matrices and the latent roots are distinct. By deleting from (1) in a symmetrical manner all except ν of the variables and equations we get a system

$$(\mathbf{C} - \lambda \mathbf{D})\mathbf{y} = \mathbf{0}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

while the effect of reintroducing one more variable and equation is to give

$$\left[\left(\frac{\mathbf{C}}{\mathbf{c}'} \middle| \frac{\mathbf{c}}{\gamma} \right) - \lambda \left(\frac{\mathbf{D}}{\mathbf{d}'} \middle| \frac{\mathbf{d}}{\delta} \right) \right] \left(\frac{\mathbf{y}}{\eta} \right) = \mathbf{0}. \quad . \quad . \quad . \quad . \quad (5)$$

In these equations \mathbf{C} and \mathbf{D} are symmetrical matrices of the ν th order, \mathbf{c} and \mathbf{d} are ν -rowed column matrices, and γ and δ are scalars. The escalator consists of a number of steps, in each of which the known latent roots $\lambda_r^{[\nu]}$ and normalized root-vectors $\mathbf{y}_r^{[\nu]}$ of (4) determine the latent roots $\lambda_s^{[\nu+1]}$ and normalized root-vectors $\left(\frac{\mathbf{y}_s^{[\nu+1]}}{\eta_s}\right)$ of (5): (4) is readily solved for $\nu=1$; so that repetition of these steps finally leads to a solution of (5) for $\nu+1=n$, i. e. a solution of (1).

Let $\mathbf{Y}=(\mathbf{y}_1^{[\nu]}, \mathbf{y}_2^{[\nu]}, \dots, \mathbf{y}_\nu^{[\nu]})$ be the matrix of normalized root-vectors of (4). Then from the standard theory of quadratic forms

$$\mathbf{D}\mathbf{Y}\mathbf{Y}'=\mathbf{Y}\mathbf{Y}'\mathbf{D}=\mathbf{Y}'\mathbf{D}\mathbf{Y}=\mathbf{I}. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Write (5) as

$$(\mathbf{C}-\lambda\mathbf{D})\mathbf{y}+(\mathbf{c}-\lambda\mathbf{d})\eta=\mathbf{0} \quad . \quad . \quad . \quad . \quad . \quad (7)$$

and

$$(\mathbf{c}'-\lambda\mathbf{d}')\mathbf{y}+(\gamma-\lambda\delta)\eta=0. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Let

$$P_r=(\mathbf{c}'-\lambda\mathbf{d}')\mathbf{y}_r^{[\nu]}. \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Transpose (7), postmultiply by $\mathbf{y}_r^{[\nu]}$, and substitute from (9)

$$\mathbf{y}'(\mathbf{C}-\lambda\mathbf{D})\mathbf{y}_r^{[\nu]}+P_r\eta=0. \quad . \quad . \quad . \quad . \quad . \quad (10)$$

By definition, from (4),

$$(\mathbf{C}-\lambda_r^{[\nu]}\mathbf{D})\mathbf{y}_r^{[\nu]}=\mathbf{0}.$$

Premultiply by \mathbf{y}' , and subtract from (10)

$$(\lambda_r^{[\nu]}-\lambda)\mathbf{y}'\mathbf{D}\mathbf{y}_r^{[\nu]}+P_r\eta=0. \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Hence

$$\mathbf{y}'\mathbf{D}\mathbf{Y}=-\left(\frac{P_1\eta}{\lambda_1^{[\nu]}-\lambda}, \frac{P_2\eta}{\lambda_2^{[\nu]}-\lambda}, \dots, \frac{P_\nu\eta}{\lambda_\nu^{[\nu]}-\lambda}\right).$$

Postmultiply by \mathbf{Y}' and use (6)

$$\mathbf{y}'=\mathbf{y}'\mathbf{D}\mathbf{Y}\mathbf{Y}'=-\left(\frac{P_1\eta}{\lambda_1^{[\nu]}-\lambda}, \dots, \frac{P_\nu\eta}{\lambda_\nu^{[\nu]}-\lambda}\right)\{\mathbf{y}_1^{[\nu]}, \dots, \mathbf{y}_\nu^{[\nu]}\}=-\eta\sum_{r=1}^{\nu}\frac{P_r\mathbf{y}_r^{[\nu]}}{\lambda_r^{[\nu]}-\lambda}.$$

So, transposing

$$\mathbf{y}=-\eta\sum_{r=1}^{\nu}\frac{P_r\mathbf{y}_r^{[\nu]}}{\lambda_r^{[\nu]}-\lambda}. \quad . \quad . \quad . \quad . \quad . \quad (12)$$

Combining (8) and (12) and substituting from (9)

$$(\gamma-\lambda\delta)=-\frac{1}{\eta}(\mathbf{c}'-\lambda\mathbf{d}')\mathbf{y}=(\mathbf{c}'-\lambda\mathbf{d}')\sum_{r=1}^{\nu}\frac{P_r\mathbf{y}_r^{[\nu]}}{\lambda_r^{[\nu]}-\lambda}=\sum_{r=1}^{\nu}\frac{P_r^2}{\lambda_r^{[\nu]}-\lambda}.$$

Hence

$$\phi_{[\nu]}(\lambda)\equiv\sum_{r=1}^{\nu}\frac{P_r^2}{\lambda_r^{[\nu]}-\lambda}-(\gamma-\lambda\delta)=0 \quad . \quad . \quad . \quad . \quad . \quad (13)$$

is the characteristic equation of (5), and it has roots $\lambda_s^{[v+1]}$. Now

$$\begin{aligned}\phi'_{[v]}(\lambda) &\equiv \frac{\partial}{\partial \lambda} \phi_{[v]}(\lambda) = \sum_{r=1}^v \left\{ \frac{2P_r}{\lambda_{[v]} - \lambda} \cdot \frac{\partial P_r}{\partial \lambda} + \frac{P_r^2}{(\lambda_{[v]} - \lambda)^2} \right\} + \delta \\ &= \sum_{r=1}^v \left\{ \frac{2P_r}{\lambda_{[v]} - \lambda} (-\mathbf{d}' \mathbf{y}_r^{[v]}) + \frac{P_r^2}{(\lambda_{[v]} - \lambda)^2} \right\} + \delta, \text{ from (9)} \\ &= \sum_{r=1}^v \left\{ \frac{2\mathbf{y}' \mathbf{D} \mathbf{y}_r^{[v]}}{\eta} \mathbf{d}' \mathbf{y}_r^{[v]} + \left(\frac{\mathbf{y}' \mathbf{D} \mathbf{y}_r^{[v]}}{\eta} \right)^2 \right\} + \delta, \text{ from (11)} \\ &= \frac{2\mathbf{y}' \mathbf{D}}{\eta} \left\{ \sum_{r=1}^v \mathbf{y}_r^{[v]} \mathbf{y}_r^{[v]'} \right\} \mathbf{d} + \frac{\mathbf{y}' \mathbf{D}}{\eta^2} \left\{ \sum_{r=1}^v \mathbf{y}_r^{[v]} \mathbf{y}_r^{[v]'} \right\} \mathbf{D} \mathbf{y} + \delta, \text{ transposing} \\ &\quad \text{scalar terms} \\ &= \frac{2\mathbf{y}' \mathbf{D}}{\eta} \mathbf{Y} \mathbf{Y}' \mathbf{d} + \frac{\mathbf{y}' \mathbf{D}}{\eta^2} \mathbf{Y} \mathbf{Y}' \mathbf{D} \mathbf{y} + \delta \\ &= \frac{1}{\eta^2} \{ 2\mathbf{y}' \mathbf{d} \eta + \mathbf{y}' \mathbf{D} \mathbf{y} + \eta \delta \eta \}, \text{ by (6)} \\ &= \frac{1}{\eta^2} (\mathbf{y}' | \eta) \left(\frac{\mathbf{D} | \mathbf{d}}{\delta} \right) \left(\frac{\mathbf{y}}{\eta} \right). \quad \dots \dots \dots (14)\end{aligned}$$

To satisfy (5) replace λ , \mathbf{y} by $\lambda_s^{[v+1]}$, $\mathbf{y}_s^{[v+1]}$ and write η_s for η . Then since $\left(\frac{\mathbf{y}_s^{[v+1]}}{\eta_s} \right)$ is normalized for (5), (14) and (12) give

$$\eta_s^2 = 1 / \phi'_{[v]}(\lambda_s^{[v+1]}), \quad \dots \dots \dots (15)$$

$$\mathbf{y}_s^{[v+1]} = -\eta_s \sum_{r=1}^v \frac{P_{rs} \mathbf{y}_r^{[v]}}{\lambda_{[v]} - \lambda_s^{[v+1]}}, \quad \dots \dots \dots (16)$$

where P_{rs} is P_r with $\lambda_s^{[v+1]}$ written for λ in (9). (15) and (16) give the required root-vectors associated with $\lambda_s^{[v+1]}$. From the partial fraction nature of (13) we see that the roots $\lambda_r^{[v]}$ separate the roots $\lambda_s^{[v+1]}$; and hence, by backward induction from the hypothesis that the latent roots of (1) are simple, it follows that the latent roots of (13) are simple. Hence (15) is not nugatory. This completes the proof.

In the case of the collineatory form (2), $P_r = \mathbf{c}' \mathbf{y}_r^{[v]}$ is independent of λ ; and the solution of the successive equations (13) for $v=1, 2, \dots, n-1$ becomes much simplified. Nevertheless the labour of having to find all the roots of each such equation is still very considerable; and there is also the set of matrix multiplications necessary at each stage for the formation of P_r and $\mathbf{y}_s^{[v+1]}$. Morris and Head state:—"The escalator process does not involve the evaluation of determinants; neither does it involve iteration. The only successive approximation device used is the Newtonian method for approximating to the roots of an equation." The first part of this statement is misleading in so far as the set of matrix multiplications

mentioned above implicitly evaluates the respective determinants : and the second part of the statement seems a contradiction in terms ; for the Newtonian method is certainly an iterative one, just as the evaluation of a square root (*e. g.* in numerically solving the equation $x^2=2$) is an iterative process. Also, the Newtonian method will apply to (13) only if all the roots are real.

Samuelson's Method.

This method (Samuelson 1942) only determines the coefficients of the reduced characteristic function. It springs from the observation that a homogeneous set of simultaneous differential equations

$$D\mathbf{x}=\mathbf{T}\mathbf{x}$$

(where D denotes the differential operator $d/d\theta$) implies †, for any individual component x_1 of \mathbf{x} , that

$$\phi_1(D)x_1=0$$

(where ϕ_1 is the reduced characteristic function of \mathbf{T}) is the result of eliminating the remaining components of \mathbf{x} from $D\mathbf{x}=\mathbf{T}\mathbf{x}$. Writing $\mathbf{x}^{(i)}=D^i\mathbf{x}$, successive differentiations yield

$$\left. \begin{array}{rcl} -\mathbf{I}\mathbf{x}^{(1)}+\mathbf{T}\mathbf{x} & = & 0 \\ -\mathbf{I}\mathbf{x}^{(2)}+\mathbf{T}\mathbf{x}^{(1)} & = & 0 \\ \dots\dots\dots & & \\ -\mathbf{I}\mathbf{x}^{(n)}+\mathbf{T}\mathbf{x}^{(n-1)} & = & 0 \end{array} \right\},$$

i. e. a set of n^2 equations in $n(n+1)$ unknowns

$$\{\mathbf{x}^{(n)} \mid \mathbf{x}^{(n-1)} \mid \dots \mid \mathbf{x}^{(1)} \mid \mathbf{x}\}.$$

If we eliminate the n^2-1 variables other than $x_1, x_1^{(1)}, \dots, x_1^{(n)}$, we may expect to get $\phi_1(D)x_1=0$. So write

$$\mathbf{T}=\left(\begin{array}{c|c} t_{11} & \mathbf{r}' \\ \hline \mathbf{s} & \mathbf{Q} \end{array}\right);$$

† Strictly this is not true, although it contains the pith of Samuelson's approach. The equation $D\mathbf{x}=\mathbf{T}\mathbf{x}$ only implies $f_1(D)x_1=0$ where f_1 is a factor of ϕ_1 . Almost all linear combinations $\xi=\sum_{i=1}^n a_i x_i$ are such that $D\mathbf{x}=\mathbf{T}\mathbf{x}$ implies $f_\xi(D)\xi=0$, where $f_\xi=\phi_1$ (apart from a trivial numerical factor) : so that, taking $\xi=x_1$, we should be unlucky if f_1 turned out to be a polynomial of lower degree than ϕ_1 . However, unless the reduced characteristic function ϕ_1 coincides with the characteristic function ϕ , there is no linear combination ξ , such that $D\mathbf{x}=\mathbf{T}\mathbf{x}$ implies $\phi(D)\xi=0$. The coincidence of ϕ_1 with ϕ occurs if, and only if, the Segre characteristic of \mathbf{T} contains no internal brackets. Samuelson (1942) stated that this method yielded the coefficients of ϕ ; so that his statement is still further from the truth than the modified account I give of it in the text above.

and let \mathbf{J} be the unit matrix of $(n-1)$ rows, $\mathbf{0}$ be a null square matrix of $(n-1)$ rows, \mathbf{o} be a null column of $(n-1)$ rows, and

$$\mathbf{W} = \left[\begin{array}{cccccc|cccccc} -\mathbf{J} & \mathbf{Q} & \mathbf{0} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{o} & -\mathbf{s} & \mathbf{o} & \dots & \mathbf{o} & \mathbf{o} \\ \mathbf{0} & -\mathbf{J} & \mathbf{Q} & \dots & \mathbf{0} & \mathbf{0} & \mathbf{o} & \mathbf{o} & -\mathbf{s} & \dots & \mathbf{o} & \mathbf{o} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \dots & -\mathbf{J} & \mathbf{Q} & \mathbf{o} & \mathbf{o} & \mathbf{o} & \dots & \mathbf{o} & -\mathbf{s} \\ \mathbf{o}' & \mathbf{o}' & \mathbf{o}' & \dots & \mathbf{o}' & \mathbf{r}' & 0 & 0 & 0 & \dots & 1 & -t_{11} \\ \mathbf{o}' & \mathbf{o}' & \mathbf{o}' & \dots & \mathbf{r}' & \mathbf{o}' & 0 & 0 & 0 & \dots & -t_{11} & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ \mathbf{o}' & \mathbf{r}' & \mathbf{o}' & \dots & \mathbf{o}' & \mathbf{o}' & 1 & -t_{11} & 0 & \dots & 0 & 0 \end{array} \right],$$

having $(n-1)(n+1)$ columns before the partition and $(n+1)$ columns after it. Then an Aitken (*cf.* Aitken 1937 a) or Doolittle reduction, which annihilates the prepartition elements of \mathbf{W} , will result in a single post-partition row whose elements are the coefficients of ϕ_1 , because \mathbf{W} is the set of detached coefficients of a suitable rearrangement of our n^2 equations.

The reduction of so vast an array as \mathbf{W} may at first sight appear a formidable undertaking: Samuelson remarks that this is not really so, for most of the elements of \mathbf{W} are zero; but it is not very apparent how much ingenuity would be needed to take advantage of these zeros when handling \mathbf{W} on a highly automatic computer. Samuelson concludes his paper with an interesting discussion of the "power" of a computational procedure, which he defines to be inversely proportional to the number of [non-zero] multiplications involved. He states that his method is asymptotically (for large n) more powerful than the determinantal method, and 8/9ths as powerful as the Cayley-Hamilton method with Doolittle reduction (see below); while for $n < 6$ its power exceeds the Cayley-Hamilton-Doolittle method.

Horst's Method.

Horst (1935) uses the fact that the trace of \mathbf{T}^i is s_i , the sum of the i th powers of the latent roots. Also if $\phi(\lambda) = \sum_{i=0}^n k_i \lambda^i$ is the characteristic function, Newton's identities give a triangular set of equations

$$k_n s_i + k_{n-1} s_{i-1} + \dots + k_{n-i+1} s_1 + i k_{n-i} = 0 \quad (i=1, 2, \dots, n)$$

which upon direct back-solution yield the coefficients of ϕ . The main labour of the process resides in the calculation of all the elements of $\mathbf{T}^2, \mathbf{T}^3, \dots, \mathbf{T}^{n-1}$, and the diagonal elements of \mathbf{T}^n .

The Cayley-Hamilton Method.

The Cayley-Hamilton theorem states that a matrix satisfies its own characteristic equation

$$\sum_{i=0}^n k_i \mathbf{T}^i = \mathbf{0}.$$

Hence postmultiplying by an arbitrary column vector \mathbf{u} and writing

$$\mathbf{H}=(\mathbf{u}, \mathbf{T}\mathbf{u}, \dots, \mathbf{T}^n\mathbf{u}),$$

$$\mathbf{k}=\{k_0, k_1, \dots, k_n\},$$

the coefficients of the characteristic equation are the solution of the homogeneous simultaneous equations

$$\mathbf{H}\mathbf{k}=\mathbf{0},$$

which can be reduced by either Doolittle or Aitken method provided that the latent roots are all distinct. The earliest explicit statement of this method, which I have been able to discover, is by Frazer, Duncan, and Collar (1938). The new method to be propounded in this article is an extension of the Cayley-Hamilton method, in two respects. Firstly we shall replace the characteristic function of \mathbf{T} by its minimum functions; and consequently we shall be able to deal with multiple roots as well as distinct simple roots. Secondly we shall provide additional steps by which the root-vectors can be determined at the same time as the latent roots.

§3. PROPOSED NEW METHOD.

Theoretical Treatment.

The language of the following discussion is mainly geometrical: Todd (1947) [Chap. 5] and Turnbull and Aitken (1945) [Chaps. 4, 5, 6] provide the relevant theoretical background. But first an outline statement of the underlying ideas may help to clarify the exposition. If \mathbf{T}_1 is the rational canonical form of \mathbf{T} , we may write (2) in the form

$$(\mathbf{T}_1 - \lambda \mathbf{I})\mathbf{z} = \mathbf{0}, \quad \dots \quad (17)$$

$$\mathbf{x} = \mathbf{G}\mathbf{z}, \quad \dots \quad (18)$$

where $\mathbf{T} = \mathbf{G}\mathbf{T}_1\mathbf{G}^{-1}$, and \mathbf{G} is the matrix whose columns are the successive independent object and image points of a complete principal sequence in S_{n-1} . (17) is readily solved for its latent roots and root-vectors; and then (18) provides the root-vectors in the original coordinate system. This is standard theory; but the objection to it as a procedure for practical computing lies in the unnecessary labour of determining a complete principal sequence. We shall overcome this objection by modifying \mathbf{G} to a form \mathbf{H} , in which the complete principal sequence is replaced by a set of arbitrary vectors chosen at random from S_{n-1} .

We shall place no restrictions whatsoever on the nature of \mathbf{T} ; it need not be symmetrical, it may contain complex elements, it may be singular or even null, its characteristic equation may have real or complex roots, the roots may be simple or multiple, and any multiple root may lead either to a unique repeated root-vector or to an invariant space.

We shall find it convenient to associate polynomials with vectors: more precisely, a polynomial $\phi(\lambda) \equiv \sum_{j=0}^s k_j \lambda^j$ will be associated with the

$(s+1)$ -rowed column vector $\mathbf{k} = \{k_0, k_1, \dots, k_s\}$. Also if λ_r is a root of $\phi(\lambda)=0$, the quotient $\phi(\lambda)/(\lambda-\lambda_r)$ will exist as a polynomial, and will be associated with the s -rowed column vector, which we write ${}^r\mathbf{k}$. We shall refer to \mathbf{k} and ${}^r\mathbf{k}$ respectively as the vector of ϕ and the quotient-vector of ϕ with respect to λ_r .

Let a set of non-null vectors $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_l$ be a sequence of points in the (augmented) projective space S_{n-1} satisfying the conditions

- (i) \mathbf{u}_1 is a point of maximum index, p_1 ,
- (ii) \mathbf{u}_i is a point of maximum relative index, p_i , with respect to $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{i-1}$ for $i=2, 3, \dots, l$,
- (iii) $p_i > 0$ for $i=2, 3, \dots, l$; subject to which l is a maximum.

[In particular, if $p_1=n$, the sequence will consist of a single point \mathbf{u}_1 : this case includes, inter alia, the case when all the latent roots of \mathbf{T} are distinct.] We may observe that $p_1 > 0$, by definition as an index (as opposed to a relative index); that l is finite, since the p 's are positive integers whose sum is n ; and that the conditions are consistent, for they are certainly satisfied by any complete principal sequence of S_{n-1} . We shall show later, in fact, that if l vectors are drawn at random from S_{n-1} , there is zero probability that they fail to satisfy these three requirements.

Let

$$\left. \begin{aligned} \mathbf{H}^{(i)} &= (\mathbf{u}_i, \mathbf{T}\mathbf{u}_i, \dots, \mathbf{T}^{p_i-1}\mathbf{u}_i) \\ \mathbf{h}^{(i)} &= \mathbf{T}^{p_i}\mathbf{u}_i \end{aligned} \right\} (i=1, 2, \dots, l). \quad (19)$$

Since \mathbf{u}_i has relative index p_i , the equations

$$(\mathbf{H}^{(1)} | \mathbf{H}^{(2)} | \dots | \mathbf{H}^{(i-1)} | \mathbf{H}^{(i)}, \mathbf{h}^{(i)}) \{ \mathbf{k}_i^{(1)} | \mathbf{k}_i^{(2)} | \dots | \mathbf{k}_i^{(i-1)} | \mathbf{k}_i \} = \mathbf{0}. \quad (20)$$

have a unique solution (apart from a scalar multiple), the component $\mathbf{k}_i^{(j)}$ of the partitioned column matrix $\{ \}$ having p_i rows and the component \mathbf{k}_i having (p_i+1) rows. The column matrices are the vectors of the polynomials $\phi_i^{(j)}$ and ϕ_i in the relation

$$\left\{ \sum_{j=1}^{i-1} \phi_i^{(j)}(\mathbf{T})\mathbf{u}_j \right\} + \phi_i(\mathbf{T})\mathbf{u}_i = \mathbf{0}. \quad (21)$$

Now it is known that ϕ_i divides $\phi_i^{(j)}$ with quotient $Q_i^{(j)}$, say, and that

$$\mathbf{v}_i = \mathbf{u}_i + \sum_{j=1}^{i-1} Q_i^{(j)}(\mathbf{T})\mathbf{u}_j \quad (22)$$

defines a complete principal sequence $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_l$. Substituting in (21) we get

$$\phi_i(\mathbf{T})\mathbf{v}_i = \mathbf{0},$$

where it is easy to see that the degree of ϕ_i is a minimum and the index of \mathbf{v}_i is p_i . Hence $\phi_1, \phi_2, \dots, \phi_l$ are a complete set of minimum functions of \mathbf{T} . Let λ_r be any root of $\phi_i(\lambda)=0$: then it is also a root of $\phi_i^{(j)}(\lambda)=0$.

Hence we can write (20) in the form

$$(\mathbf{T} - \lambda_r \mathbf{I}) \mathbf{x}_r^{(i)} = \mathbf{0}, \quad (23)$$

where

$$\mathbf{x}_r^{(i)} = (\mathbf{H}^{(1)} | \mathbf{H}^{(2)} | \dots | \mathbf{H}^{(i-1)} | \mathbf{H}^{(i)}) \{ r\mathbf{k}_i^{(1)}, 0 | r\mathbf{k}_i^{(2)}, 0 | \dots | r\mathbf{k}_i^{(i-1)}, 0 | r\mathbf{k}_i \}. \quad (24)$$

Now $\mathbf{x}_r^{(i)}$ is non-null; for otherwise (24) would violate the condition that the relative index of \mathbf{u}_i is p_i . It follows from (23) that the vectors $\mathbf{x}_r^{(i)}$, given by (24), are the base elements of the invariant space associated with λ_r . (If λ_r is a root of ϕ_1 , but not of $\phi_2, \phi_3, \dots, \phi_l$, then this invariant space degenerates to a root-vector.)

We finally prove that there is zero probability that a set of l vectors drawn at random from S_{n-1} will not satisfy conditions (i), (ii), and (iii). Suppose firstly that $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{\alpha-1}$ do satisfy these three conditions ($\alpha \leq l$); and that a new vector \mathbf{u}_α , selected at random from S_{n-1} , is added to them. Since $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{\alpha-1}$ satisfy the conditions, (22) for $i=1, 2, \dots, \alpha-1$ defines the first $\alpha-1$ points $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{\alpha-1}$ of a complete principal sequence, whose remaining points are, say, $\mathbf{v}_\alpha^*, \mathbf{v}_{\alpha+1}^*, \dots, \mathbf{v}_l^*$. Then

$$\xi_i^{(j)} = \begin{cases} \mathbf{T}^j \mathbf{v}_i & (i=1, 2, \dots, \alpha-1; j=0, 1, \dots, p_i-1), \\ \mathbf{T}^j \mathbf{v}_i^* & (i=\alpha, \alpha+1, \dots, l; j=0, 1, \dots, p_i-1) \end{cases} .$$

is a minimal complete base set of S_{n-1} ; so we may write

$$\mathbf{u}_\alpha = \sum_{i,j} a_{\alpha i}^{(j)} \xi_i^{(j)}.$$

Suppose that the relative index of \mathbf{u}_α with respect to $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{\alpha-1}$ is less than p_α : then there exist polynomials ψ_α and $\psi_\alpha^{(j)}$ such that

$$\psi_\alpha(\mathbf{T}) \mathbf{u}_\alpha + \sum_{j=1}^{\alpha-1} \psi_\alpha^{(j)}(\mathbf{T}) \mathbf{u}_j = \mathbf{0}$$

and

$$\deg(\psi_\alpha) < p_\alpha.$$

Hence

$$\psi_\alpha(\mathbf{T}) \sum_{i=\alpha}^l \sum_{j=0}^{p_i-1} a_{\alpha i}^{(j)} \mathbf{T}^j \mathbf{v}_i^* = \text{terms involving only } \mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{\alpha-1} \text{ and} \\ \mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{\alpha-1}. \quad (25)$$

But we may use (22) to eliminate from the right-hand side of (25) the terms in $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{\alpha-1}$. Then in (25) the left-hand side consists entirely of terms in $\mathbf{v}_\alpha^*, \mathbf{v}_{\alpha+1}^*, \dots, \mathbf{v}_l^*$ and the right-hand side of terms in $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{\alpha-1}$. Further, these points being of a complete principal sequence, their characteristic subspaces do not intersect. It follows that

every individual term in this version of (25) vanishes ; and in particular for the term in \mathbf{v}_α^*

$$\psi_\alpha(\mathbf{T}) \sum_{j=0}^{p_\alpha-1} a_{\alpha\alpha}^{(j)} \mathbf{T}^j \mathbf{v}_\alpha^* = \mathbf{0},$$

whence

$$\psi_\alpha(\mathbf{T}) \sum_{j=0}^{p_\alpha-1} a_{\alpha\alpha}^{(j)} \mathbf{T}^j \equiv \mathbf{0} \pmod{\phi_\alpha(\mathbf{T})}.$$

Suppose that (including multiplicities where necessary) λ_r ($r=1, 2, \dots, p_\alpha$) are the roots of $\phi_\alpha(\lambda)=0$. Then since $\deg(\psi_\alpha) < p_\alpha$, at least one of the relations

$$\pi_r \equiv \sum_{j=0}^{p_\alpha-1} a_{\alpha\alpha}^{(j)} \lambda_r^j = 0 \quad (r=1, 2, \dots, p_\alpha)$$

holds. But, since $\{a_{\alpha i}^{(j)}\}$ is a point coordinate set with respect to the complete base set $\{\xi_i^{(j)}\}$, $\pi_r=0$ is the equation of a fixed prime in S_{n-1} , the prime coordinates being either zero or powers of a latent root of \mathbf{T} . The set of points in any prime of S_{n-1} is of zero $(n-1)$ -dimensional

measure ; so, since \mathbf{u}_α belongs to the finite set of fixed primes $\sum_{r=1}^{p_\alpha} \pi_r$, it also belongs to a set of zero $(n-1)$ -dimensional measure. There is thus zero conditional probability that a random vector \mathbf{u}_α has, with respect to

- $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{\alpha-1}$, a relative index less than p_α , when it is known that $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{\alpha-1}$ satisfy conditions (i), (ii), and (iii). This is true for all values of $\alpha \geq 2$. A few changes of wording in the above argument show that there is zero probability that \mathbf{u}_1 , chosen at random from S_{n-1} , is not a point of maximum index. Consequently, upon successive combination of these zero probabilities for $\alpha=1, 2, \dots, l$, we find that there is zero probability that a random set $\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_l$ fails to satisfy conditions (i), (ii), and (iii).

The importance of this probability result is that, while we cannot assert that the method proposed here will certainly work, we can at least be assured that there is zero probability of failure. As an analogy, the Newtonian method of approximating to the real cube root of 2 by means of the recurrence relation

$$x_{i+1} = \frac{2(1+x_i^3)}{3x_i^2}; \quad x_0 \text{ real}$$

will only fail if we start from some number x_0 which belongs to the sequence of real numbers defined by

$$y_{i+1}^3 - \frac{3}{2} y_{i+1}^2 y_i + 1 = 0; \quad y_0 = 0.$$

Almost none of the real numbers belong to this sequence ; so there is zero probability of this method failing when x_0 is selected at random from the field of real numbers.

Practical Application.

A numerical example will illustrate the application of the proposed method. Consider (1) with

$$\mathbf{A} = \begin{bmatrix} -10 & -14 & -8 & -6 & -2 & -8 \\ 31 & 44 & 21 & 13 & 5 & 25 \\ -12 & -18 & -15 & -10 & -3 & -9 \\ -12 & -15 & 51 & -8 & -3 & -14 \\ 18 & 27 & 17 & 13 & 4 & 14 \\ 0 & 4 & 6 & 5 & 1 & -1 \end{bmatrix},$$

and

$$\mathbf{B} = \begin{bmatrix} 6 & 9 & 1 & -2 & 0 & 5 \\ -2 & -4 & -7 & -4 & -1 & -1 \\ 8 & 13 & 67 & 4 & 1 & 2 \\ -2 & -1 & 1 & 1 & 0 & -2 \\ 5 & 7 & 4 & 3 & 1 & 4 \\ 0 & -4 & -6 & -5 & -1 & 1 \end{bmatrix}.$$

As we shall see in due course, five of the six latent roots in this example have equal moduli, there is a triple root possessing only a unique root-vector, and there is a double root which leads to an invariant space of dimensionality 1. Clearly this example would prove awkward if treated by ordinary methods.

The first stage (see Table I.) consists of the formation of the quotient matrix $\mathbf{T} = \mathbf{B}^{-1}\mathbf{A}$. This is done by Aitken's method (Aitken 1932, 1937 a)† and calls for few comments. The figures in heavy type are pivotal elements; and the italic figures in the extreme right-hand column are the complements of row sums, which provide a line-by-line running check of the computation. To provide checks in the second stage of the work, we

† This, and the subsequent reduction in Table II., might also have been done by the Doolittle method. Personally, I prefer Aitken's method when the matrix is not symmetrical, and the square root method (Laderman 1948) when the matrix is symmetrical: but this is largely a matter of taste. See also Hoel (1941). It is worth remarking that even when \mathbf{A} and \mathbf{B} are both symmetrical, \mathbf{T} will not in general be symmetrical. The abbreviated Doolittle method (Dwyer 1941) is inappropriate when the matrix to be reduced is not symmetrical. Crout's method (Crout 1941) is effectively an abbreviated Doolittle method, valid in the general case of a non-symmetrical matrix; and it also bears a close affinity to the square root method, since the Crout auxiliary matrix is a diagonally bordered triangular set. We could certainly use Crout's method in these reductions, and thereby achieve quite a considerable compression of Tables I. and II. as far as the number of entries to be written down is concerned; but I have eschewed this method in the present example, because I wished to avoid the use of decimals or fractions. Should the divisor matrix be singular, Aitken's reduction will (at any rate in theory) detect the fact; whereupon we should interchange \mathbf{A} and \mathbf{B} and start afresh.

TABLE I.—First stage—Reduction to collineatory form by matrix predivision.

A

29	—8
—120	25
—28	—9
4	—14
—117	14
0	—1

6	9	1	—2	0	5
—2	—4	—7	—4	—1	—1
8	13	67	4	1	2
—2	—1	1	1	0	—2
5	7	4	3	1	4
0	—4	—6	—5	—1	1

B

—662	134	26	66	110	—8	—14	—10	—14	—8	—6	—2	—8
—400	10	—2	—12	—26	21	44	31	44	21	13	5	25
82	—100	—22	—60	290	—15	—18	—12	—15	—15	—10	—3	—9
—847	124	34	108	142	51	—15	—12	—15	51	—8	—3	—14
0	—6	6	30	36	17	27	18	27	17	13	4	14
1062	—144	—24	—54	—84	6	4	0	4	6	5	1	—1
1242	—168	—30	—72	—510								
516	—57	—21	—75	—87								
—2648	542	98	234	404								
86022	—11640	—2058	—4896	—31098								
23541	—2427	—1083	—4074	—4587								
—178180	34954	6278	14922	25572								
—785493	99771	23259	68202	250251								
415770	—149442	—27264	—64746	183474								

53345	0	—2929	—7215	11525	—22775	—16951	—16951	—22775	11525	—7215	—2929	—15501	0
—7118	0	391	962	—1548	2975	2265	2265	2975	—1548	962	391	2072	0
—3497	0	192	473	—756	1460	1111	1111	1460	—756	473	192	1016	0
10401	0	—588	—1439	2605	—4448	—3400	—3400	—4448	2605	—1439	—588	—3132	0
—48887	0	2760	6759	—12155	20877	15954	15954	20877	—12155	6759	2760	14691	0
—46336	0	2537	6255	—9852	19301	14680	14680	19301	—9852	6255	2537	13414	0
	0	—2363	—5795	10181	—17890	—13659	—13659	—17890	10181	—5795	—2363	—12560	0

T*

add a final column of zeros to \mathbf{T} and a final row such that the sum of each column in the resultant $(n+1) \times (n+1)$ matrix, \mathbf{T}^* , is zero. These figures are also shown in italics.

In the second stage (see Table II.) we produce the vectors of the functions $\phi_i^{(j)}$ and ϕ_i . Instead of starting with an n -rowed arbitrary column vector \mathbf{u}_1 , we start with an $(n+1)$ -rowed vector \mathbf{u}_1^* , the sum of whose elements is zero, but which is otherwise arbitrary. Since the fifth column of \mathbf{T}^* is numerically the smallest, we take for convenience

$$\mathbf{u}_1^* = \{0, 0, 0, 0, 1, 0, -1\}.$$

The heads of the next five columns of Table II. are $\mathbf{T}^*\mathbf{u}_1^*$, $\mathbf{T}^{*2}\mathbf{u}_1^*$, ..., $\mathbf{T}^{*5}\mathbf{u}_1^*$. Actually we do not produce them in this order. We first calculate $\mathbf{T}^*\mathbf{u}_1^*$, and then (according to Aitken's method) reduce it by the pivotal element in the left-hand column. This gives the second set of figures in the second column. Then we multiply $\mathbf{T}^*\mathbf{u}_1^*$ by \mathbf{T}^* to give the head of the third column, which we thereupon reduce by the pivotal elements in the first and second columns. And so on : such that in general for the head of the $(i+1)$ th column we get $\mathbf{T}^{*i}\mathbf{u}_1^*$ by premultiplying $\mathbf{T}^{*i-1}\mathbf{u}_1^*$ by \mathbf{T}^* ; whereupon condensation by the pivotal elements of the preceding columns yields the remainder of the column. Heavy type indicates pivotal elements ; and italic type marks the elements derived from the additional row of \mathbf{T}^* and the additional element of \mathbf{u}_1^* . The computation thus consists of a number of subcolumns of elements, each subcolumn comprising n or fewer elements in normal or heavy type and one final element in italic type ; and it is easy to see that every such subcolumn must have zero sum (for the transformations are confined to a prime of S_n^* defined by this property). So the italic figures in Table II. provide running checks of every step in the work. In the reduction we may choose the pivotal elements in any convenient row ; but to provide convenient routine checks in the back solution stage (described presently) the italic rows should not contain pivotal elements.

Now either this process terminates in the production of n pivotal elements, or else there comes a stage when the last set of reduced elements in a column are all zero, so that there is no non-zero pivotal element available for the reduction of succeeding columns. Thus in Table II, if we reduce the final set of the sixth column

$$-1224$$

$$-504$$

$$1728$$

by the corresponding entries in the previous column

$$-306$$

$$-126$$

$$432$$

we get

$$0$$

$$0.$$

TABLE II.—Second stage—Determination of the functional vectors.

$H^{(1)*}$	$h^{(1)*}$	$H^{(2)*}$	$h^{(2)*}$
0	—2929	—14903	—43731
0	391	1992	5845
0	192	977	2867
0	—588	—3040	—8919
1	2760	14256	41822
0	2537	12885	37810
—1	—2363	—12167	—35694
—2929	—2929	—14903	—43731
391	391	1992	5845
192	192	977	2867
—588	—588	—3040	—8919
2537	2537	12885	37810
397	2089	2089	6128
	—7495	—7495	—21184
	—257	—1091	—3172
	141196	409923	1116144
	68746	200057	544924
	—202190	—587705	—1600606
		983	3090
		—16625	—54976
		—8053	—26636
		23745	78522
		—306	—1224
		—126	—504
		432	1728
		306	—432
		—1224	1728
		1224	—1728
		612	—864
		—1530	2160
		612	—864
		k_1	
		—16951	—16951
		2265	2265
		1111	1111
		—3400	—3400
		14680	14680
		2295	2295
		—6344	—6344
		473	473
		—8588	—8588
		6967	6967
		7492	7492
		—1767	—1767
		22048	22048
		11713	11713
		—31994	—31994
		34263	34263
		12846	12846
		—47109	—47109
		414	414
		—414	—414
		414	414
		51566	51566
		—170200	—170200
		86526	86526
		163990	163990
		—144302	—144302

We stop manipulating \mathbf{u}_1^* at this stage, and start afresh with a new arbitrary vector \mathbf{u}_2^*

$$\mathbf{u}_2^* = \{1, 0, 0, 0, 0, 0, -1\}.$$

This we reduce by every pivotal element that has so far occurred, and so on for $\mathbf{T}^*\mathbf{u}_2^*$, $\mathbf{T}^{*2}\mathbf{u}_2^*$, When we are next unable to find a non-zero pivotal element, we start afresh with another arbitrary vector \mathbf{u}_3^* , and so on. Eventually we must attain a stage in which there are n pivotal elements. In the present example \mathbf{u}_1^* and \mathbf{u}_2^* suffice to reach this final stage; but in the general case we shall have got the array

$$(\mathbf{H}^{(1)*}, \mathbf{h}^{(1)*} | \mathbf{H}^{(2)*}, \mathbf{h}^{(2)*} | \dots | \mathbf{H}^{(l)*}, \mathbf{h}^{(l)*})$$

together with its Aitken reductions to triangular form; so that by ordinary back solution we immediately obtain the solutions of (20) for all the values $i=1, 2, \dots, l$. Moreover, these back solutions result not only from the rows containing pivotal elements, but also from the rows of italic figures; and the two solutions should agree (apart from a scalar multiple). This checks the back solutions. In the example under consideration we find

$$\mathbf{k}_1 = \{612, -1530, 612, 1224, -1224, 306\},$$

$$\{\mathbf{k}_2^{(1)} | \mathbf{k}_2\} = \{-144302, 163990, 86526, -170200, 51566 | 414, 414\},$$

and the corresponding figures in italics in Table II. are the checks just mentioned. For convenience, removing arbitrary scalar multiples, we get

$$\mathbf{k}_1 = \{2, -5, 2, 4, -4, 1\} \quad (\text{on dividing by } 306),$$

$$\{\mathbf{k}_2^{(1)} | \mathbf{k}_2\} = \{-3137, 3565, 1881, -3700, 1121 | 9, 9\} \quad (\text{on dividing by } 46).$$

Our next task is to solve the characteristic equation

$$\phi(\lambda) \equiv \phi_1(\lambda)\phi_2(\lambda) \dots \phi_l(\lambda) = 0.$$

In this, the work is considerably simplified by the knowledge that, for every value of i , the roots of $\phi_{i+1}(\lambda)=0$ are also roots of $\phi_i(\lambda)=0$ with at least the same degree of multiplicity. So we may solve in turn $\phi_l(\lambda)=0$, $\phi_{l-1}(\lambda)=0$,, $\phi_1(\lambda)=0$. As we shall see presently, we may in the special case when the latent roots are real adopt an iterative method of solution which yields the quotient-vectors as a by-product of determining the latent roots. For the moment, however, suppose that we know the value of λ_r , a root of $\phi_i(\lambda)=0$. Then, since the following set of equations are identical with the long division process for dividing $\phi_i^{(j)}(\lambda)$ by $\lambda - \lambda_r$,

$$\left. \begin{aligned} r k_{i, p_j-2}^{(j)} &= k_{i, p_j-1}^{(j)} \\ r k_{i, p_j-3}^{(j)} &= k_{i, p_j-2}^{(j)} + \lambda_r r k_{i, p_j-2}^{(j)} \\ &\dots \dots \dots \\ r k_{i, \alpha-1}^{(j)} &= k_{i, \alpha}^{(j)} + \lambda_r r k_{i, \alpha}^{(j)} \\ &\dots \dots \dots \\ r k_{i, 0}^{(j)} &= k_{i, 1}^{(j)} + \lambda_r r k_{i, 1}^{(j)} \\ 0 &= \phi_i^{(j)}(\lambda_r) = k_{i, 0}^{(j)} + \lambda_r r k_{i, 0}^{(j)} \end{aligned} \right\}, \dots \dots \dots (26)$$

we have a method of evaluating $r\mathbf{k}_i^{(j)}$ which embodies the check inherent in the last equation of (26). By writing p_i+1 for p_j and dropping the superfix (j) in (26) we likewise obtain $r\mathbf{k}_i$. We may also note that the check just mentioned is not only a check of the computed elements of $r\mathbf{k}_i$ and $r\mathbf{k}_i^{(j)}$, but is as well a check on the correctness of the latent root λ_r . Now (26) is merely a sequence of operations in which a set of numbers c_β ($\beta=1, 2, \dots$) are fed in, and from which we produce results $d_\beta=c_\beta+\lambda d_{\beta-1}$ with $d_0=0$ and λ fixed throughout. On the other hand, using the property of the last equation of (26), we see that if $\lambda=\lambda_r$ and we feed in serially the complete set

$$\{\mathbf{k}_i^{(1)} | \mathbf{k}_i^{(2)} | \dots | \mathbf{k}_i^{(i-1)} | \mathbf{k}_i\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (27)$$

in reverse order as though it were a single set, then the result will be the reverse order of

$$\{0, r\mathbf{k}_i^{(1)} | 0, r\mathbf{k}_i^{(2)} | \dots | 0, r\mathbf{k}_i^{(i-1)} | 0, r\mathbf{k}_i\}. \quad . \quad . \quad . \quad (28)$$

Rearranging the partitions and omitting the first zero, (28) is

$$\{r\mathbf{k}_i^{(1)}, 0 | r\mathbf{k}_i^{(2)}, 0 | \dots | r\mathbf{k}_i^{(i-1)}, 0 | r\mathbf{k}_i\}, \quad . \quad . \quad . \quad . \quad (29)$$

which is required in (24). This observation is, of course, trivial when we are dealing with simple calculating machines; but it becomes of some interest for a highly automatic computer—*e.g.* with punched-card analysis we can treat (27) as a single pack of cards fed to a multiplying punch of which the output is (29). We can indeed go further; for, if λ_r is a latent root leading to an invariant space of dimensionality $\mu-1$, it will be a root of $\phi_i(\lambda)=0$, ($i=1, 2, \dots, \mu$). Hence, by feeding in all the packs (27) for $i=1, 2, \dots, \mu$ as a single super-pack, we shall obtain the complete set of quotient-vectors (29) for this invariant space.

Returning to the question of solving $\phi_i(\lambda)=0$ for real roots, we see that, even if λ is not a root of this equation, (26) will yield the value of $\phi_i(\lambda)$. A similar set of equations (in which we write $\alpha k_{i,\alpha}$ for $k_{i,\alpha}$ and omit the last equation) will yield $\phi'_i(\lambda) \equiv \partial \phi_i(\lambda) / \partial \lambda$. In effect, this is the process suggested by Mack (1944) †. We have the Newtonian approximation that, if λ_0 is an approximate root, then

$$\lambda_0 - \phi_i(\lambda_0) / \phi'_i(\lambda_0) \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

is a better approximation, while

$$\lambda_0 - \frac{2\phi_i(\lambda_0)}{\phi'_i(\lambda_0) + \phi'_i\{\lambda_0 - (\phi_i(\lambda_0) / \phi'_i(\lambda_0))\}}$$

is still better. [If the second term of the denominator of the correction part of this second expression is found by linear interpolation in ϕ'_i , the error involved is normally of the third order of the correction itself.]

† Porter and Mack (1949) have extended this method to complex roots.

Whether or not the use of ϕ'_i is worth while depends a good deal on the value of n : for large n it will usually be preferable to work entirely in terms of ϕ_i (obtained by (26)) and its divided differences.

In general when ϕ_i has complex zeros, we must abandon the Newtonian method in favour of, say, the Graeffe method: but once these zeros are estimated, we can use (26) to check them and produce the quotient-vectors (29).

The final stage of the computation is the substitution of (29) in (24). All the quantities involved (including the italic check sums in **H**) are already known; so that we immediately obtain the root-vectors $\mathbf{x}_r^{(i)}$ and a set of relevant check sums. Where, for fixed r , there exist several distinct $\mathbf{x}_r^{(i)}$, these are the base of the invariant space associated with λ_r .

The work for the last stages of the computation appears in Table III.

TABLE III.

Third stage—Calculation of Quotient-vectors and Root-vectors.

\mathbf{k}_1	\mathbf{rk}_1			$\{\mathbf{k}_2^{(1)} \mathbf{k}_2\}$	$\{\mathbf{rk}_2^{(1)}, 0 \mathbf{rk}_2\}$
	$\lambda=-1$	$\lambda=1$	$\lambda=2$		$\lambda=-1$
	Simple root	Triple root	Simple root		Simple root
1				9	
-4	1	1	1	9	9
4	-5	-3	-2	1121	0
2	9	1	0	-3700	1121
-5	-7	3	2	1881	-4821
2	2	-2	-1	3565	6702
				-3137	-3137
Check $[\phi]$	0	0	0		0
\mathbf{x}_r^*	-4851 654 318 -1089 5091 4146 -4269	-2379 318 156 -487 2283 2056 -1947	-28278 3780 1854 -5778 27090 24444 -23112		3194763 -430705 -209426 716865 -3351331 -2730601 2810435

We enter the functional vectors (27) in reverse order; and applying (26) we get the quotient-vectors (29) in reverse order. Lastly, from Tables II. and III., we have

$$\mathbf{x}_r^{(1)*} = \mathbf{H}^{(1)*} \mathbf{rk}_1 \quad \text{and} \quad \mathbf{x}_r^{(2)*} = (\mathbf{H}^{(1)*} | \mathbf{H}^{(2)*}) \{\mathbf{rk}_2^{(1)}, 0 | \mathbf{rk}_2\};$$

and these products give the \mathbf{x}_r^* shown in Table III., together with their italic check sums. We can, if we wish, simplify the numerical values of

the root-vectors by taking out arbitrary scalar factors : thus we divide $\mathbf{x}_{\lambda=2}^{(1)}$ by 18 ; and for the invariant space arising when $\lambda=-1$ we take one base element as

$$\frac{1434213}{25047} \mathbf{x}_{\lambda=-1}^{(1)} + \frac{2178}{25047} \mathbf{x}_{\lambda=-1}^{(2)},$$

and the other as

$$\frac{237644}{25047} \mathbf{x}_{\lambda=-1}^{(1)} + \frac{363}{25047} \mathbf{x}_{\lambda=-1}^{(2)}.$$

This gives the final results :—

(i) Double latent root $\lambda=-1$ leads to an invariant line whose base is

$$\{ 33, -4, -2, -21, 95, -40 \},$$

$$\{ 275, -37, -18, 57, -267, -237 \}.$$

(ii) Triple latent root $\lambda=1$ leads to a unique root-vector

$$\{ -2379, 318, 156, -487, 2283, 2056 \}.$$

(iii) Simple latent root $\lambda=2$ leads to a unique root-vector

$$\{ -1571, 210, 103, -321, 1505, 1358 \}.$$

For a final check we substitute these values in (1).

Computing Machinery.

For $n \leq 10$, roughly, desk calculators are quite sufficient to effect a solution by this proposed new method. I worked the foregoing example ($n=6$) on an electric desk calculator with automatic multiplication (Fridén ST 10) quite comfortably in one day, despite focusing the greater part of my attention on the layout of the computation rather than its execution. The total number of multiplications is, however, roughly $5n^3$; so that for $10 < n \leq 15$ the method is going beyond the scope of desk calculators (the case $n=15$ would probably occupy three computers on desk calculators for a week), and for $n > 15$ it becomes more or less unmanageable †.

† The increase of labour results not only from the number of operations increasing as n^3 , but also from the increase in the number of digits retained during the work to override cumulative rounding-off errors, etc. (Tukerman 1941, von Neumann and Goldstine 1947). The latter authors state that it is safe if one retains an extra 8, 10, or 12 digits throughout the work in the cases $n=15$, 50, or 150 respectively ; but they admit that this stringency might be relaxed by judging the accuracy of the results according to more statistical criteria. Personally I feel that 8 guard digits is purchasing safety at a very high and indeed most dangerous price, and I should be quite satisfied with 5 guard digits for $n=15$. The risks originating from an excess of guard digits are, however, less threatening the less is the human element an agent of the computation ; so with punched-card analysis one might tranquilly accept a greater factor of safety from extra guard digits.

However, punched-card machinery will readily extend the range of n . The methods of pivotal condensation used in the first two stages of the work need modifying slightly. As previously described they follow the normal Aiken procedure of dividing the cross-products by previous pivotal elements in order to keep down the size of entries; and this demands about n^3 divisions. On punched-card equipment one should avoid division wherever possible. It is merely necessary to determine the reciprocals of pivotal elements (and, in the second stage, of the corresponding leading italic check figures) and to multiply pivotal (and check) rows by these reciprocals before making the back solutions. Altogether this calls for $3n$ divisions, which can be worked out on an auxiliary desk calculator. Some further divisions arise from (30) in determining the latent roots; but this number is small—if each root requires half a dozen iterations, one has to make a total of $6n$ divisions (or less if there are multiple roots). In all, the auxiliary desk calculator performs about $9n$ divisions, while the punched-card equipment is entirely freed from divisions and may concentrate on the main burden of $5n^3$ multiplications and associated additions. The technique for reductions of the type in question on standard punched-card equipment (Comrie 1933, Eckert 1940) or the more powerful perforation relay computers (Aiken *et al.* 1946, Eckert 1948) has already received attention (Hartley 1946, Mitchell 1948). On the basis of Mitchell's results ($59\frac{1}{2}$ hours to invert a matrix of order 38), we may surmise that about $(2n/15)^3$ hours are needed on a relay computer for the present more complex problem, which requires two inversions and the additional steps of the third stage. Thus, working a 40-hour week, a relay computer should solve the case $n=50$ in about 7 weeks.

With present or projected facilities (Aiken *et al.* 1948) there seems little hope of improving on such a performance by means of electronic equipment; for the capacity of the high-speed memory is a decisive factor. Assuming the whole calculation performed within the high-speed section, the second stage of the computation taxes the high-speed memory with a peak vocabulary of nearly $6n^2$ words apart from instructions. Consequently, if it has a high-speed memory capacity of 4000 words, the computer will only handle the cases $n \leq 25$, roughly; and for these cases the number of operations is barely sufficient to be worth undertaking on an electronic scale. Alternatively, assuming part of the storage in the low-speed memory, one faces the difficulty of repeatedly selecting words from large arrays of about n^2 elements and returning words thereto. The time thus absorbed is comparable with the reading time of a relay computer; so that little advantage arises from using electronic equipment. This state of affairs illustrates the interesting point that certain computational methods are too concise for an electronic computer. Electronic computers thrive when very many operations manage comparatively few data: whereas the problem under discussion has quite a lot of data (**A** and **B** possess $2n^2$ individual elements) and quite a lot of answers (n latent roots and n^2 elements in the root-matrix), yet our solution resolves the problem by relatively few operations (some $5n^3$ multiplications).

When highly automatic machinery is available, the Bernoulli-Aitken method (Aitken 1926) might prove suitable for evaluating complex roots of the minimum functions $\phi_i(\lambda)$. To use this method for latent roots is not really to follow the circular path of trying to solve problem (A) by reducing it to a solution of problem (B) and then attacking (B) by reducing it to a solution of (A): for, when n is large, the alternation from one standpoint to the other and then back again does simplify the structure considerably, and may perhaps unify the programming of the computer because of the similarity between (A) and (B).

Comparison with Current Methods.

I believe that the proposed new method is an improvement on current methods.

The determinantal method involves $O(n^4)$ operations, as opposed to $O(n^3)$ operations in the new method.

We can make a direct comparison with the Duncan-Collar method for the example used in the text by employing the figures $\mathbf{u}_1, \mathbf{T}\mathbf{u}_1, \dots, \mathbf{T}^5\mathbf{u}_1$ in Table II. This comparison is fair, because the simple root $\lambda=2$ is the root of greatest magnitude and is indeed twice as large as any other root. We find

m	$f(m)$	$\psi(m)$	$\psi_2(m)/\delta^2\psi(m)$
0	1	2760.0	—
1	2760	8394.384	5023.766
2	23168499	5.11214 2	5.11243 6
3	118440668	5.11243 62	5.11214 25
4	605520363	2.93413 32556	2.93413 32986
5	1776677434	2.93413 32986	2.93413 32566
6	5213008420	2.51366 77144	2.51366 86920
7	13103770960	2.51366 86920	2.51366 77144
8	32938538809	2.27012 64570	2.27012 64581
9	74774648406	2.27012 64581	—
10	169747907739	—	—

The last two columns of this table are evidently tending to 2, as they should; but the approximation given by the Duncan-Collar method is not very impressive for the few terms that suffice for an exact solution by the new method. The δ^2 -process, illustrated by the last column merely permutes the order of the sequence and does not improve the estimate afforded by ψ ; but this is not surprising because Aitken's sufficient condition, that there should be three dominant simple roots, is not satisfied. Note also that the foregoing table applies merely to one of the six roots.

The weakest link of the new method lies in having to solve the minimum functions $\phi_i(\lambda)$. This seems unavoidable, because even in the simplest

case when the data is rational the final answers are in general algebraic numbers from an n th degree equation. Nevertheless, the new method has an air of being as economical as possible †; because it only requires the determination of a single set of n roots, and it also simplifies this set still further by distributing it between the minimum functions whenever such a distribution exists. Contrast this with the Morris escalator, in which one has to solve algebraic equations of degrees 2, 3, 4, . . . , $n-1$, n each for all their roots (*i. e.* in all, a set of $\frac{1}{2}n(n+1)-1$ roots).

Samuelson's method only solves part of the problem, and is even then slightly less powerful than the new method.

Horst's method also provides only a partial solution; and the number of operations is $O(n^4)$.

The Cayley-Hamilton method is another partial solution. The new method refines it by completing the solution.

Simplifications in Special Cases.

If **B** is symmetrical, the first stage (Table I.) of the new method can be shortened by using the square root method of reduction or the abbreviated Doolittle method. Similarly we can shorten the second stage (Table II.), if **T** is symmetrical.

If we merely want the latent roots and root-vectors of a matrix, we can omit the whole of the first stage (Table I.).

If we only want the values of distinct latent roots and a root-vector for each such value, we may omit everything in the second stage (Table II.) except $\mathbf{H}^{(1)*}$, $\mathbf{h}^{(1)*}$, and \mathbf{k}_1 .

If we do not require root-vectors, we omit the second half of the third stage (Table III.).

If we only want a particular set of latent roots and their root-vectors (*e. g.* the smallest and the largest latent roots), we evaluate the only corresponding columns of Table III. This simplification applies in particular for the largest latent root alone; but in this particular case we may prefer the Duncan-Collar method.

† In a certain sense one may perhaps consider that the new method belongs to that class of methods which have optimum computational simplicity: for concepts of algebraic fields show that in theory the solution must involve at least two $n \times n$ matrix reductions, the solution of one n th degree algebraic equation, and one $n \times n$ matrix multiplication. The new method involves exactly this. But I regard any such notion as highly tentative: for what holds in theory may not apply in practice. There is difficulty in defining the idea of optimal simplicity in the face of a knowledge that our answers are always correct merely to such and such a number of decimal places: and although it would be nice to set up a hierarchy of computational methods arranged according to their simplicity, one must remember for instance that the recent discovery of an elementary proof of the prime number theorem has at any rate ruffled the Hardy-Littlewood concept of the depth of a mathematical theorem.

REFERENCES.

- AIKEN, H. H. *et al.*, 1946, "Manual of Operation for the Automatic Sequence Controlled Calculator", *Ann. of the Computation Laboratory of Harvard University*—I. (London: Oxford University Press); 1948, "Proceedings of a Symposium on Large-Scale Digital Calculating Machinery", *Ann. of the Computation Laboratory of Harvard University*.—XVI. (London: Oxford University Press).
- AITKEN, A. C., 1926, "On Bernoulli's numerical solution of algebraic equations", *Proc. Roy. Soc. Edin.*, **46**, 289–305; 1932, "On the evaluation of determinants, the formation of their adjugates and the practical solution of simultaneous linear equations", *Proc. Edin. Math. Soc.*, [2], **3**, 207–219; 1937 a, "Studies in practical mathematics.—I. The evaluation, with applications, of a certain triple product matrix", *Proc. Roy. Soc. Edin.*, **57**, 172–181; 1937 b, "Studies in practical mathematics.—II. The evaluation of the latent roots and latent vectors of a matrix", *Ibid.*, **57**, 269–304.
- ALT, F. L., 1946, "Multiplication of matrices", *Math. Tables and other Aids to Computation*, **2**, 12–13.
- BRODETSKY, S., and SMEAL, G., 1924, "On Graeffe's method for complex roots of algebraic equations", *Proc. Camb. Phil. Soc.*, **22**, 83–87.
- COMRIE, L. J., 1933, *The Hollerith and Powers Tabulating Machines* (reprint of Newmarch lectures: printed for private circulation); 1936, "Inverse interpolation and scientific applications of the National accounting machine", *J. Roy. Statist. Soc. Suppl.*, **3**, 87–114; 1937, *Interpolation and Allied Tables* (reprinted from Nautical Almanac, 1937) (London: H. M. Stationery Office).
- CORNOCK, A. F., and HUGHES, J. M., 1943, "The evaluation of the complex roots of algebraic equations", *Phil. Mag.* [7], **34**, 314–320.
- CROUT, P. D., 1941, "A short method for evaluating determinants and solving systems of linear simultaneous equations with real or complex coefficients", *Trans. Amer. Inst. Elec. Eng.*, **60**, 1235–1240.
- DOOLITTLE, M. H., 1878, "Method employed in the solution of normal equations and the adjustment of triangulation", *U.S. Coast and Geodetic Survey Report*, 115–120.
- DUNCAN, W. J., and COLLAR, A. R., 1934, "A method for the solution of oscillation problems by matrices", *Phil. Mag.* [7], **17**, 865–909.
- DWYER, P. S., 1941, "The Doolittle technique", *Ann. Math. Statist.*, **12**, 449–485.
- ECKERT, W. J., 1940, *Punched Card Methods in Scientific Computation*, Thomas J. Watson Astronomical Computing Bureau, Columbia University.; 1948, "I.B.M. pluggable sequence relay calculator", *Math. Tables and other Aids to Computation*, **3**, 149–161.
- FRAZER, R. A., and DUNCAN, W. J., 1929, "On the numerical solution of equations with complex roots", *Proc. Roy. Soc. A*, **125**, 68–82.
- FRAZER, R. A., DUNCAN, W. J., and COLLAR, A. R., 1938, *Elementary Matrices and some Applications to Dynamics and Differential Equations* (Cambridge University Press).
- HARTLEY, H. O., 1946, "The application of some commercial calculating machines to certain statistical calculations", *J. Roy. Statist. Soc. Suppl.*, **8**, 154–183.
- HARTREE, D. R., 1948, "Experimental arithmetic", *Eureka*, **10**, 13–18.
- HOEL, P. G., 1941, "On methods of solving normal equations", *Ann. Math. Statist.*, **12**, 354–359.
- HORST, P., 1935, "A method of determining the coefficients of a characteristic equation", *Ann. Math. Statist.*, **6**, 83–84.

- HOTELLING, H., 1943 a, "Some new methods in matrix calculation", *Ann. Math. Statist.*, **14**, 1-34; 1943 b, "Further points on matrix calculation and simultaneous equations", *Ibid.*, **14**, 440-441.
- LADERMAN, J., 1948, "The square root method for solving simultaneous linear equations", *Math. Tables and other Aids to Computation*, **3**, 13-16.
- MACK, C., 1944, *R.R.D.E. Research Report No. 249* (restricted publication).
- MITCHELL, H. F., 1948, "Inversion of a matrix of order 38", *Math. Tables and other Aids to Computation*, **3**, 161-166.
- MORRIS, J., and HEAD, J. W., 1944, "The escalator process for the solution of Lagrangian frequency equations", *Phil. Mag.* [7], **35**, 735-759.
- PORTER, A., and MACK, C., 1949, "New methods for the numerical solution of algebraic equations", *Phil. Mag.* [7], **40**, 578-585.
- SAMUELSON, P. A., 1942, "A method of determining explicitly the coefficients of the characteristic equation", *Ann. Math. Statist.*, **13**, 424-429.
- TODD, J. A., 1947, *Projective and Analytical Geometry* (London: Pitman and Sons).
- TUKERMAN, L. B., 1941, "On the mathematically significant figures in the solution of simultaneous linear equations", *Ann. Math. Statist.*, **12**, 307-316.
- TURNBULL, H. W., and AITKEN, A. C., 1945, *An Introduction to the Theory of Canonical Matrices*, 2nd ed. (London and Glasgow: Blackie & Sons).
- VON NEUMANN, J., and GOLDSTINE, H. H., 1947, "Numerical inverting of matrices of high order", *Amer. Math. Soc. Bull.*, **53**, 1021-1099.
- WEATHERBURN, C. E., 1917, "A plea for a more general use of vector analysis in applied mathematics", *Math. Gaz.*, **9**, 2-5.
- WEDDERBURN, J. H. M., 1934, *Lectures on Matrices* (New York: Amer. Math. Soc. Colloq. Publ., **17**).
- WHITTAKER, E. T., and ROBINSON, G., 1940, *The Calculus of Observations*, 3rd ed. (London and Glasgow: Blackie & Sons).

LXXXVI. *The Emission of Short-range Alpha Particles from Light Elements under Proton Bombardment.—I. Experimental Method and the Reaction $^{10}\text{B}(p\alpha)^7\text{Be}$.*

By W. E. BURCHAM* and JOAN M. FREEMAN†.

[Received May 26, 1949.]

SUMMARY.

A method for studying the emission of short-range alpha particles in the bombardment of light nuclei by protons is described. The alpha particles have been resolved from the scattered protons by a 90° focusing magnet and detected by a fluorescent screen and photo-electric multiplier.

The results of an investigation of the $^{10}\text{B}(p\alpha)$ reaction by this method are presented. The Q-value of the reaction has been found to be $1.11 \text{ MeV.} \pm 0.06$, and the cross-section to be $1.2 \times 10^{-26} \text{ cm.}^2$ at a proton energy of 530 keV. A target of separated ^{10}B isotope was used.

* Selwyn College, Cambridge; † Newnham College, Cambridge.†

‡ Communicated by O. R. Frisch.

§ 1. INTRODUCTION,

THE bombardment of light nuclei by protons of energy up to 1 MeV. leads in many cases to capture of the proton and subsequent emission of gamma radiation. Sharp resonances for the emission of this capture radiation are often observed, and the conditions for the occurrence of these resonance phenomena have been discussed by Fowler, Lauritsen and Lauritsen (1948). The sharpest levels are found when the only process which can compete with the emission of gamma radiation is re-emission of the incident proton, and this process then determines the observed width of the level. If the emission of alpha particles from the level is both energetically possible and allowed by selection rules, the width will be further increased; if the alpha particles are emitted with energy sufficient to take them over the potential barrier the width may be as much as 1 MeV. No sharp resonances are then observed, either for alpha particles or gamma-ray emission, and the variation of the yield of these processes with proton energy is determined mainly by the penetrability of the potential barrier of the bombarded nucleus for the incident proton.

If, however, the alpha particles are emitted with low energy, so that they have to penetrate a wide barrier, the alpha-particle width is reduced by a penetrability factor. The level width may then again be chiefly determined by proton re-emission, and sharp resonances in the yield of both alpha particles and gamma radiation can be observed. Such resonances will only be found if there are levels of the compound nucleus in the accessible range of excitation energy. If there are not, the yield of the reactions will again be chiefly determined by barrier penetrabilities, since levels remote from the region of investigation will not contribute a rapidly varying disintegration probability. The occurrence of resonance phenomena in ($p\alpha$) reactions therefore requires not only a low energy release, but also a high excitation of the compound nucleus.

The cases in which these conditions are fulfilled are shown, together with other ($p\alpha$) reactions in the light elements, in Table I. Column 2 gives the energy release in the reaction calculated from the masses given by Bethe (1947), and column 3 the excitation energy of the compound nucleus formed in the reaction. It is clear that resonance effects may be expected in cases such as $^{15}\text{N}(p\alpha)^{12}\text{C}$, $^{19}\text{F}(p\alpha, \gamma)^{16}\text{O}^*$, $^{23}\text{Na}(p\alpha)^{20}\text{Ne}$, and $^{27}\text{Al}(p\alpha)^{24}\text{Mg}$ where the excitation of the compound nucleus is fairly high and the energy of the emitted alpha particle is low. On the other hand, for reactions such as $^6\text{Li}(p\alpha)^3\text{He}$, $^9\text{Be}(p\alpha)^6\text{Li}$, $^{10}\text{B}(p\alpha)^7\text{Be}$ and $^{17}\text{O}(p\alpha)^{14}\text{N}$, the excitation of the compound nucleus is relatively low, and it is unlikely that many resonance levels can affect the yield of the reaction for proton energies up to 1 MeV. Columns 4 and 5 of Table I. give barrier heights B_p and B_α for the target and residual nuclei in the reaction; these have been calculated using the formula given by Bethe (1937) with the nuclear radii proposed by Amaldi and Cacciapuoti (1947). In column 6 is the product $P_p P_\alpha$ of the barrier penetrabilities for the particles taking part

in the reaction for a proton energy of 1 MeV. and an angle of observation of 90° ; the figure given measures, apart from the influence of selection rules, the relative difficulty of detecting the reaction.

The observation of $(p\alpha)$ reactions in which the energy release is small obviously becomes difficult for target nuclei of atomic number much greater than that of aluminium. The investigation of such reactions is also complicated experimentally by the fact that the range of the alpha particles emitted will often be less than that of the protons scattered from the target. For targets of the type frequently used in transmutation experiments, in which a thin layer of material is deposited on a water-cooled copper disk, many more protons than alpha particles will be observed in a given solid angle. If the target can be deposited on a thin foil of a light metal such as beryllium, the number of scattered protons may be so

TABLE I.

$(p\alpha)$ reactions in the light elements.

Reaction	Q (MeV.) (Masses)	Excitation of compound nucleus (MeV.)	B _p (MeV.)	B _a (MeV.)	P _p P _a
${}^6\text{Li}(p\alpha){}^3\text{He}$	3.90	5.5	1.5	4.6	0.2
${}^7\text{Li}(p\alpha){}^4\text{He}$	17.3	17.2	1.4	3.7	0.9
${}^9\text{Be}(p\alpha){}^6\text{Li}$	2.11	6.5	1.6	4.2	0.08
${}^{10}\text{B}(p\alpha){}^7\text{Be}$	1.15	8.7	2.0	5.0	0.002
${}^{11}\text{B}(p\alpha){}^8\text{Be}$	8.60	15.9	1.9	4.6	0.5
${}^{15}\text{N}(p\alpha){}^{12}\text{C}$	4.92	12.1	2.4	5.5	0.1
${}^{17}\text{O}(p\alpha){}^{14}\text{N}$	1.13	5.7	2.6	5.9	6×10^{-5}
${}^{18}\text{O}(p\alpha){}^{15}\text{N}$	3.93	7.9	2.6	5.6	0.05
${}^{19}\text{F}(p\alpha){}^{16}\text{O}$	8.12	12.9	2.9	6.2	0.08
${}^{19}\text{F}(p\alpha, \gamma){}^{16}\text{O}^*$	1.93*	12.9	2.9	6.2	5×10^{-4}
${}^{23}\text{Na}(p\alpha){}^{20}\text{Ne}$	1.52	11.0	3.2	7.2	2×10^{-7}
${}^{27}\text{Al}(p\alpha){}^{24}\text{Mg}$	1.52	10.6	3.6	7.9	6×10^{-8}

much reduced that it is possible to distinguish the alpha particles by means of a differential ionization chamber, a proportional counter or photographic plate (Rubin 1948). Accurate measurement of the range of alpha particles of less than 1 MeV. energy is, however, not easy by these methods, especially if the particles have to pass through a mica window. In the present work the alpha particles have been resolved from the scattered protons by a 90° deflecting magnet, and have been detected by a scintillation screen and photoelectric multiplier. In this way targets deposited on ordinary thick copper supports could be used, and no thin windows were needed. The $(p\alpha)$ reactions in lithium, beryllium, boron, fluorine, sodium and aluminium have been studied by this method.

* See Burcham and Freeman 1949 a.

§ 2. EXPERIMENTAL METHOD.

(a) *The Magnetic Analyser.*

Protons scattered from a target may be resolved from transmutation alpha particles either by electrostatic or by magnetic deflection. If the focusing effects of fringing fields are neglected, both magnetic and electrostatic deflectors in their simple form give direction focusing in one plane only for particles of uniform energy, and the instruments have comparable solid angles for the same resolving power. A magnetic analyser was chosen for the experiments to be described so that alpha particles of range nearly equal to that of protons of energy 1 MeV. (the maximum available) could conveniently be studied. The corresponding alpha-particle energy is about 3 MeV., and the deflecting voltage necessary in an electrostatic analyser for particles of this energy is rather large. In addition, magnetic deflection gives resolution in momentum rather than energy per unit charge, so that if the use of an analyser is limited by protons of energy E_p scattered from a thick target, then the magnetic analyser can be used for alpha-particle energies down to E_p , while the lower limit for the electrostatic instrument is $2 E_p$. If a thin target is used, the range of alpha-particle energies over which scattered protons are detected is twice as great for the electrostatic as for the magnetic analyser.

The design of magnetic analyser adopted was similar to that described by Ringo (1940), and is based on the general theory given by Herzog (1934). The geometrical arrangement of the analyser is shown in fig. 1 (a). Particles from the source S are deflected through 90° and focused on the collector C by a magnetic field perpendicular to the plane of the diagram; AO, OB are the boundries of the field. A 90° sector has the advantage that both source and collector can be outside the magnetic field, which is convenient experimentally for the study of artificial transmutations. The geometrical properties of this type of analyser have been fully discussed (Herzog 1934, Stephens 1934, Ringo 1940); the sector-shaped field acts as a thick lens of focal length

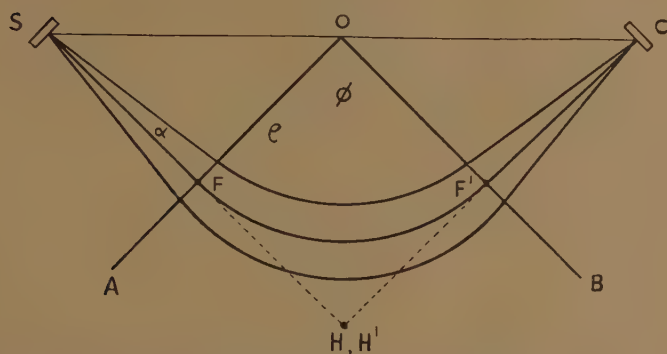
$$f = \rho / \sin \Phi,$$

where ρ is the radius of the path of the particle in the field, and Φ is the angular deflection of the particle due to the field. The principal points of the lens are at a distance

$$h = -\rho \tan \frac{1}{2} \Phi$$

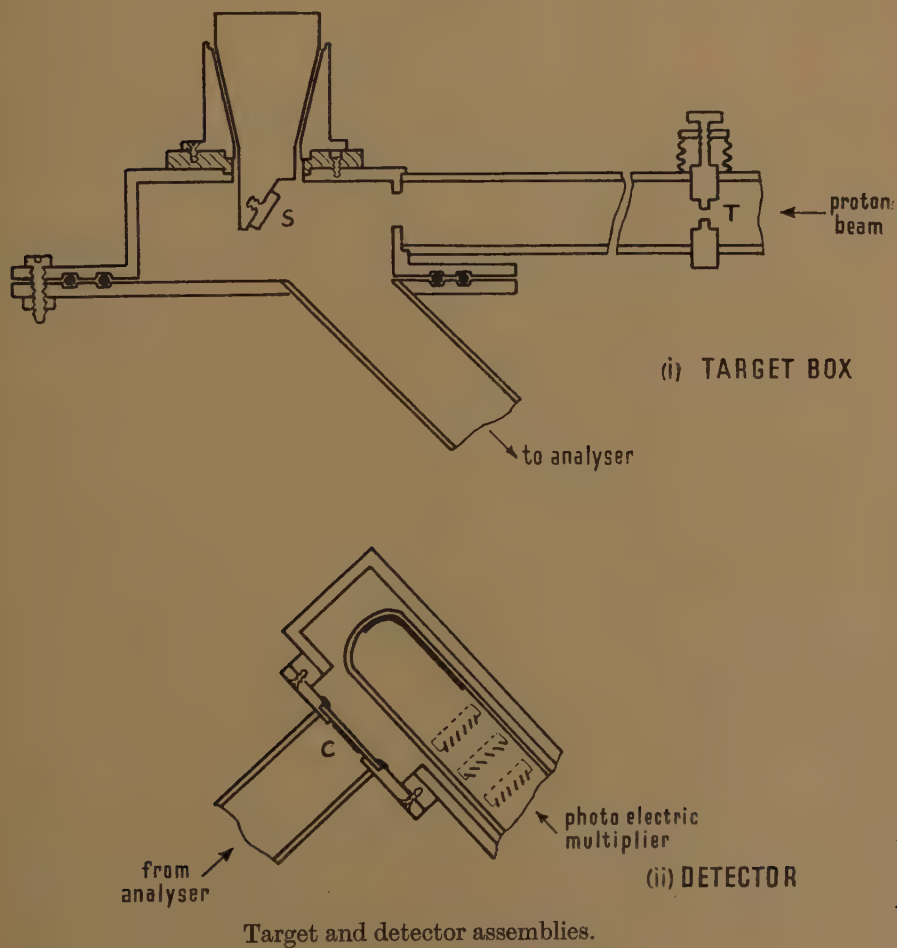
from the entrance and exit planes bounding the magnetic field. In the analyser in use, the dimensions are $\rho = 17$ cm., $\Phi = 90^\circ$, and hence $f = \rho = 17$ cm., and $h = -\rho = -17$ cm. The source and collector are placed each at a distance of $2f$ from the principal points H, H' and are thus at a distance f from the boundaries of the field. The magnification is theoretically unity and the image is inverted. The positions of the foci, principal points, source and collector are shown in fig. 1 (a).

Fig. 1 (a).



Magnetic analyser, geometrical arrangement (not to scale). $SF = FH = FO = \rho$.

Fig. 1 (b).



The resolving power of such an analyser is essentially determined by the size of the source if this is large and by the semi-angle α (see fig. 1 (a)), over which particles are accepted if the source is small. Stephens (1934) shows that for an axial point source emitting particles of uniform energy, the image width in a 90° analyser is given by

$$S = \rho \alpha^2 / \sqrt{2}$$

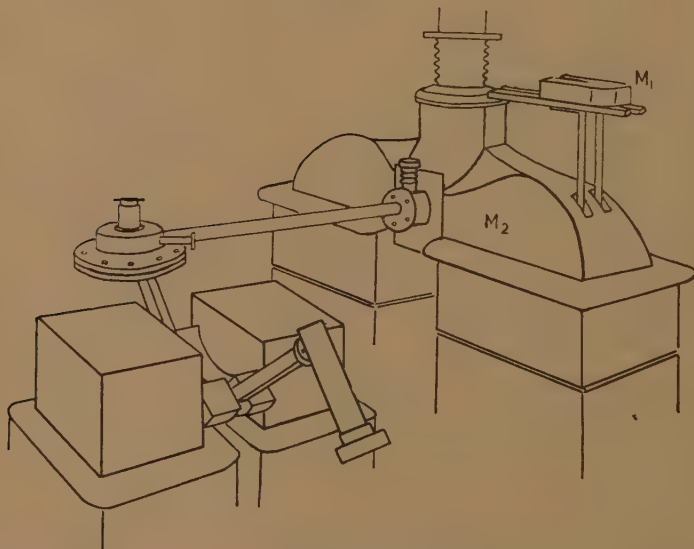
and the dispersion, for a change Δv in the particle velocity, by

$$D = \sqrt{2} \rho \Delta v / v.$$

The resolving power is defined as the value of $v/\Delta v$ for which the image is shifted through its own width and is therefore

$$R = \frac{v}{\Delta v} = \frac{\sqrt{2} \rho}{S} = \frac{2}{\alpha^2}.$$

Fig. 1 (c).



General arrangement of apparatus for study of $(p\alpha)$ reactions.

Experimentally R is usually found directly from the spread of a group of particles in momentum or energy, and may be written

$$R = \frac{v}{\Delta v} = \frac{B\rho}{\Delta(B\rho)} = \frac{2E}{\Delta E},$$

where B is the magnetic field in gauss necessary to bend particles of energy E into a circular arc of radius ρ . In the analyser in use, the semi-angle of acceptance α is defined by the width, in the plane of fig. 1 (a), of the vacuum chamber used in the instrument. This angle is $1/23$ radian, so that the theoretical resolving power for a point source is 1050. If a finite source of diameter d is used, the resolving power must be calculated

for an image shift equal to d , since the magnification is unity, and then

$$R = \frac{\sqrt{(2)\rho}}{d}$$

which, for a source of 1.75 cm. diameter, gives $R=14$. In order that this resolving power may be achieved without much sacrifice of intensity, it is desirable that the width of the collector should be approximately equal to that of the image.

The solid angle ω subtended by the analyser at the source can be calculated geometrically. With the nomenclature of fig. 1 (a), the number of particles counted when an axial point source emits N particles isotropically is

$$\frac{N\omega}{4\pi} = \frac{N}{4\pi} \frac{z}{f + \pi\rho/4} \cdot \frac{\delta}{2f + \pi\rho/2}$$

where Z is the width of the vacuum chamber in the plane of fig. 1 (a) at the centre, and δ the width of the collector perpendicular to the plane of the figure. In the analyser in use $Z=2.5$ cm. and $\delta=1.25$ cm., and the solid angle ω is therefore 1.75×10^{-3} steradians.

An ordinary laboratory magnet, with a core of cross-section 15.5 cm. \times 24 cm. was used. The poles were fitted with 90° sectors cut from a steel ring of 17 cm. central radius, and of square cross-section 5.6 cm. \times 5.6 cm. The gap was fixed at 1.7 cm., so that the vacuum chamber of the analyser was under no stress in the magnetic field. No special care was taken to reduce or to utilize the leakage field by shaping the pole pieces. The vacuum chamber of the analyser is a length of copper X-band waveguide of internal dimensions 2.5 cm. \times 1.5 cm., bent to an average radius of 17 cm. over a 90° sector and left straight for 17 cm. at each end.

(b) Experimental Tests of Magnetic Analyser.

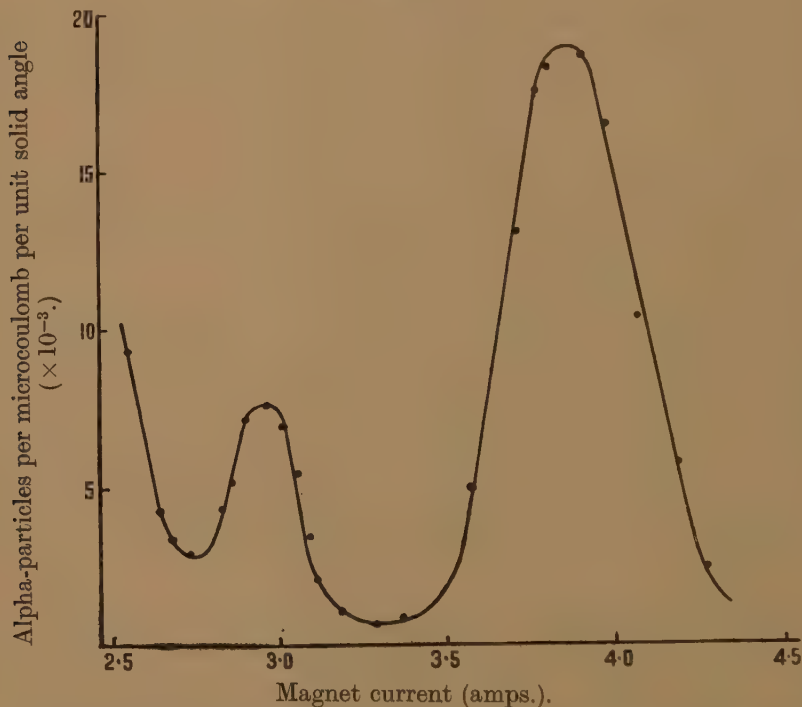
The analyser characteristics were investigated by using the short-range alpha particles from the reaction $^{19}\text{F}(p\alpha, \gamma)^{16}\text{O}^*$ (Burcham and Devons 1940). These particles are emitted from sharply defined nuclear states, and if the energy of the bombarding proton and the target thickness are suitably chosen, the spread of alpha-particle energy need not exceed the level width of about 10 keV. (Fowler, Lauritsen and Lauritsen 1948). Fig. 2 shows the number of alpha particles from a thin calcium fluoride target bombarded by protons of energy 867 keV. as a function of current in the magnet analyser coils. Such a distribution is strictly neither an energy nor a momentum spectrum, since the magnetic field is not a linear function of current, but for convenience the term spectrum will be used.

The two groups of alpha particles correspond to the gamma radiation of energy 6.13 and 6.98 MeV. observed by Walker and McDaniel (1948), and have been discussed elsewhere (Burcham and Freeman 1949 (a)). From curves of this type a figure for the resolving power is obtained by dividing the value of $B\rho$ for the peak of the curve by the width of the group, in the same units, at half maximum intensity; the result is shown in Table II.

together with similar observations for other groups studied in this work. The $B\rho$ figures shown in Table II. were derived from the magnet current readings by using the calibration curve given in fig. 8.

The table gives results for two target sizes; the agreement between observed and theoretical values is satisfactory, and, with the exception of the result for scattered protons of energy 370 keV., the resolving power seems to independent of particle energy, as would be expected.

Fig. 2.



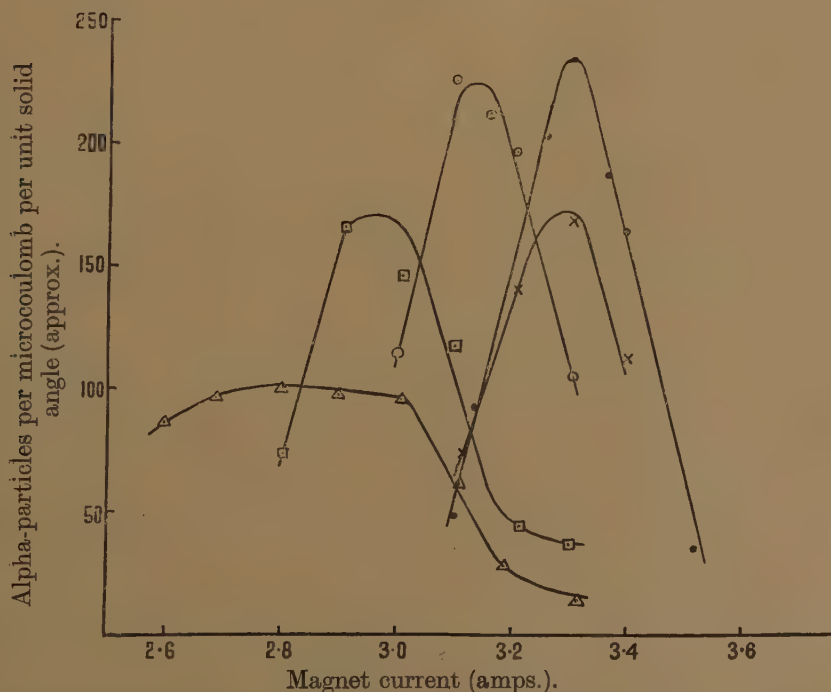
Alpha-particle groups from the reaction $^{19}\text{F}(p\alpha, \gamma)^{16}\text{O}^*$ ($E_p = 867$ keV.)
(Angle of observation = 83° .)

The resolution of the analyser is also well illustrated by the curves of fig. 3, which are typical of the behaviour of the instrument when it is used to examine resonant alpha-particle groups from thick targets. The curves are the tops of the spectra of the alpha particles from the reaction $^{27}\text{Al}(p\alpha)^{24}\text{Mg}$ (Freeman and Baxter 1948) for proton energies near the sharp resonance at 730 keV. It is clearly seen that as the proton energy increases above 730 keV., the alpha-particles are emitted from a greater depth below the surface of the target and emerge with lower energy. The intensity of the alpha-particle group decreases as the proton energy increases above 730 keV. because of increasing scattering of the alpha particles in the target, and it also decreases as the proton energy falls below 730 keV. in accordance with the width of the resonance level. The resolving power calculated from these curves is shown in Table II.

TABLE II.
Resolving Power of Magnetic Analyser.

Reaction	Particle energy (keV.)	$B\rho \times 10^{-5}$ (Gauss cm.)	$\Delta(B\rho) \times 10^{-5}$ (Gauss cm.)	R	
				1.75 cm. target and detector	1.0 cm. target and detector
$^{19}\text{F}(p\alpha_1)^{16}\text{O}^*$	2220	2.14	0.12	18	—
$^{19}\text{F}(p\alpha_2)^{16}\text{O}^*$	1610	1.82	0.11	17	—
$^6\text{Li}(p\alpha)^3\text{He}$	2010	2.04	0.105	19	—
$^9\text{Be}(p\alpha)^6\text{Li}$	1740	1.90	0.105	18	—
$^{27}\text{Al}(p\alpha)^{24}\text{Mg}$	1870	1.97	0.11	18	—
$^{19}\text{F}(p\alpha_1)^{16}\text{O}^*$	2220	2.14	0.07	—	30
$^{19}\text{F}(p\alpha_2)^{16}\text{O}^*$	1610	1.82	0.08	—	23
$^9\text{Be}(pd)^8\text{Be}$	1020	2.06	0.075	—	27
Scattered protons	370	0.88	0.05	—	17
Scattered deuterons	800	1.28	0.06	—	21
	718	1.71	0.058	—	30
	918	2.00	0.066	—	30
Theoretical				14	24.5

Fig. 3.



Alpha-particles from the reaction $^{27}\text{Al}(p\alpha)^{24}\text{Mg}$. \times 720 keV, \bullet 730 keV, \circ 750 keV, \square 770 keV, \triangle 790 keV proton energy. (Angle of observation = 90°).

The solid angle of the analyser was found by observations on the alpha particles from the reaction $^{19}\text{F}(p\alpha, \gamma)^{16}\text{O}^*$ using a proton energy of 330 keV. At this energy the range of the alpha particles is 0.95 cm., which exceeds that of the scattered protons (0.5 cm.), and they can therefore be detected by an ordinary range-measuring method. An arrangement similar to that of van Allen (1941) was used, in which the alpha particles emerged from the target chamber through a 2 mm. air equivalent mica window, traversed a variable pressure absorption cell, and were recorded by means of a scintillation screen and photoelectric multiplier at the end of the cell. A curve showing the counting rate as a function of pressure in the absorption cell was plotted, and the number of alpha particles entering the cell was found from the plateau on this curve. The alpha particle count was multiplied by a calculated solid angle factor and an experimentally determined window transmission factor to give the number of particles emitted into the whole solid angle of 4π by the target. It was assumed that the angular distribution of the alpha particles with respect to the bombarding beam was isotropic; this has been shown by van Allen (1941) and to ± 10 per cent in the present work. The same group of alpha particles was then observed after magnetic analysis, using the same detecting screen and photo-multiplier, but no absorption cell. The intensity of the bombardment was monitored in the two observations by means of a Geiger counter which responded to the gamma radiation associated with the emission of the alpha particles. In this way the solid angle ω was found to be 2×10^{-3} , which agrees well with the calculated value of 1.75×10^{-3} .

The size of the image produced at the collector screen was investigated using a small calcium fluoride target of width 2 mm. as source. The target was bombarded with protons of energy 867 keV., and the yield was monitored by a gamma-ray counter. A slit of 2 mm. width was used to scan over the collector screen near the plane of the image, and counts were taken for a number of positions of the slit. The width of the image at half maximum intensity was found to be 4.5 mm.; the value expected from the geometry of the analyser, the target thickness and the energy spread of the proton beam was 4 mm. The dispersion D was found to be 3.2 mm. for a 1 per cent change in velocity, the theoretical value for an alpha particle of 2 MeV. energy being 2.4 mm. It has been assumed from these results, which are consistent with the observed resolving power, that the image formed by the analyser is not unduly blurred, and that a perfect detector whose area exceeds that of the target will record all the particles accepted by the analyser. In computing relative intensities of different homogeneous groups of particles, the maximum ordinates of the groups have therefore been used.

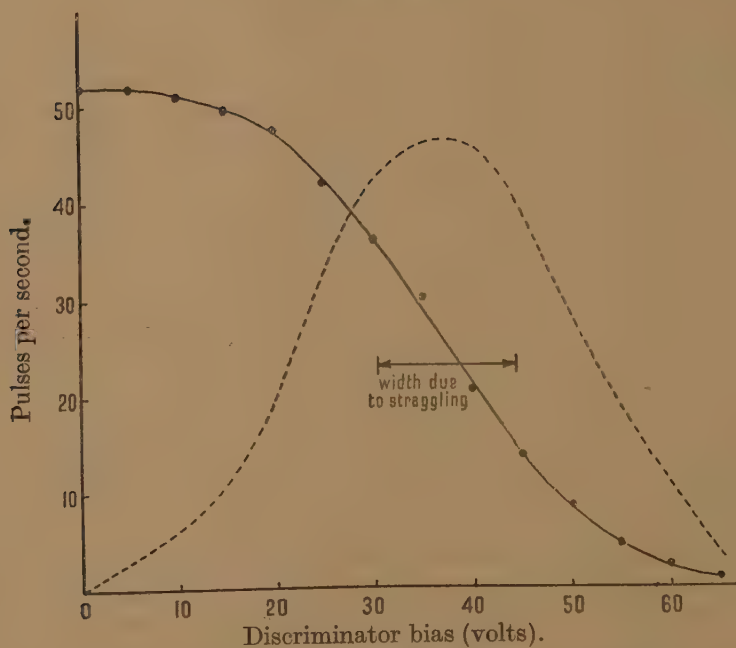
(c) Detectors.

In the first experiments a shallow air-filled ionization chamber was used and the alpha particles emerged from the vacuum chamber through a thin mica window (1.5 mm. air equivalent) mounted on a grid. Ranges.

were measured easily by drawing the chamber back, but the solid angle was limited by the window and chamber apertures to $2 \cdot 10^{-4}$, and for some alpha particles, such as those produced in the $^{10}\text{B}(p\alpha)^7\text{Be}$ disintegration, the window absorption together with the chamber penetration, exceeded the mean range of the alpha-particle group under investigation. These disadvantages were overcome by the use of the thin zinc sulphide screen and photo-multiplier to detect the alpha particles; the experimental arrangement is shown in fig. 1 (*b*). The screen was made by depositing zinc sulphide, in aqueous suspension, on a glass disk, so that after evaporation of the water a circular area of diameter 2 cm. was covered with a layer of the phosphor. The screen was sealed to the end of the vacuum chamber of the magnetic analyser, and the photo-electric multiplier (E.M.I. prototype 4588) was mounted immediately behind it, so that the photons produced when charged particles hit the screen were transmitted through the glass disk and multiplier envelope to the photo-sensitive cathode. The multiplier was housed in a steel tube which shielded it both from light and from magnetic fields. The screen thickness is of some importance, and, for alpha particles of energies of about 2 MeV., it should suffice to stop the particles. If it is much thicker the light produced by the particles is absorbed and scattered appreciably before emerging from the fluorescent layer, causing a considerable spread in pulse size for homogeneous incident particles, while if it is too thin all the alpha-particle energy is not expended in the screen and the pulse size may be too small (Broser and Kallmann 1949). The screen thickness finally chosen was 7 mgm./cm.², which gave a reasonably uniform pulse size for monokinetic alpha particles in the energy range 1 to 2 MeV. Fig. 4 (*a*) is a bias curve for the 7 mgm. screen and alpha particles of about 5 mm. residual range; the amplifier gain was reduced below normal in these observations to avoid saturation. The particles were obtained by placing absorbers in front of a thin ThC' source; the straggling thus produced contributes to the width of the pulse-size distribution derived by differentiating the bias curve. The dotted curve in fig. 4 (*a*) gives the observed distribution in pulse sizes and the horizontal line gives the estimated contribution to the width due to straggling.

The screen was also tested by observing the form of the numbers-range curve at a fixed discriminator bias when a ThC' source was gradually drawn back from the 7 mgm. screen. The curves for two bias values are shown in fig. 4 (*b*), together with a typical curve for a thicker (17 mgm./cm.²) screen. The residual mean ranges were calculated by measuring the distance from source to screen, correcting to dry air at 760 mm. pressure and 15° C., and subtracting from the mean range of ThC' alpha particles (8.57 cm.); when an inverse square law correction had been made the curves showed a flat top which was assumed to indicate an efficiency of detection of 100 per cent. In most of the experiments the bias was set between 20 and 40 volts; it is clear from fig. 4 (*b*) that under these conditions the counting efficiency is substantially constant for alpha particles of range down to about 4 mm. This uniform efficiency is very desirable

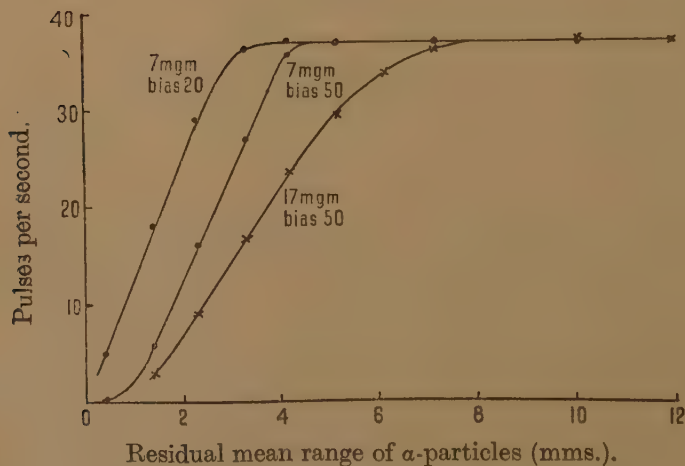
Fig. 4(a).



Characteristics of zinc sulphide screen (7 mgm./cm.²).

Bias curve —●— and pulse size distribution ---- for alpha particles of 5 mm. range.

Fig. 4(b).

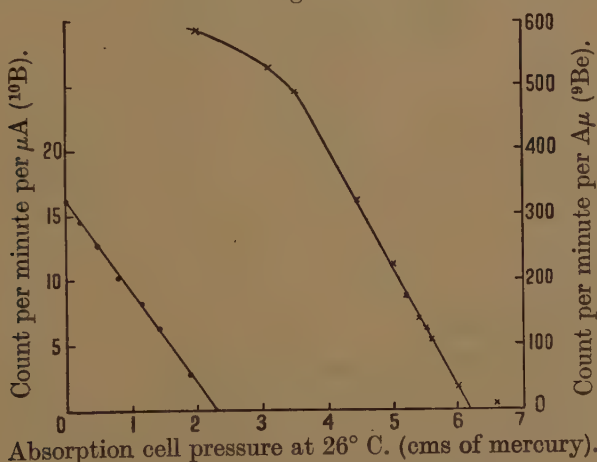


Numbers-range curve at fixed bias and normal gain.

in quantitative experiments such as the measurement of angular distributions, in which the alpha-particle energy varies with angle, and in the study of excitation functions.

Range measurements could be made with the photo-multiplier system, for sufficiently intense alpha-particle groups, by interposing a variable-pressure absorption cell between the analyser and the fluorescent screen. For this purpose the analyser was sealed with a 2 mm. window and the absorption cell, consisting of a brass cylinder 6 cm. long, was gasket-jointed to the analyser flange and terminated with the zinc-sulphide screen. The cell could be filled with air to any desired pressure, and this was read on a U-tube mercury manometer, corrections being made for temperature and humidity. The cut-off curves for alpha particles found by this means, of which fig. 5 shows typical examples, give an extrapolated range which depends on the exact mica window stopping power,

Fig. 5.



Cut-off curves for alpha-particle groups.

● $^{10}\text{B}(p\alpha)^7\text{Be}$ $E_p = 630 \text{ keV}$.

× $^9\text{Be}(p\alpha)^6\text{Li}$ $E_p = 860 \text{ keV}$.

(Angle of observation = 83° .)

the detector bias setting and the multiplier and screen efficiency; these factors were eliminated from range determinations by making comparative measurements with standard alpha-particle groups under the same conditions.

(d) Target arrangements.

The targets used were generally deposited by evaporation of the element or a salt on copper disks 1.75 cm. in diameter. These were screwed on to a water-cooled target head mounted in the target chamber at the head of the analyser as shown in fig. 1 (b). The proton beam entered the target chamber in a horizontal direction from the accelerating equipment, and the bent waveguide of the analyser emerged from the base of the target chamber at an angle of 45° with the horizontal. The analyser magnet

was mounted on a large ball-bearing joint, and the base plate of the target chamber was gasket-jointed so that the analyser could be set at different angles with the proton beam by rotating the whole assembly through an angle (β) about a vertical axis passing through the target. The actual angle θ of the axis of the analyser with the proton beam is then given by $\cos \theta = 1/\sqrt{2} \cos \beta$, and a range of angles from 48° to 134° could be covered; observations could be made in the two opposite semicircles. The target head was cut so that its surface made an angle of 22.5° with the vertical, and it was always set so that the normal to the target surface lay in the vertical plane bisecting the angle between the vertical planes passing through the proton beam and the principal section of the analyser.

Proton currents of about $30 \mu\text{A}$ could be focused on the target, and accelerating voltages of up to 1000 kV. were available.

The beam was centred on the target by a small magnet M1 mounted near the accelerating equipment; this magnet was used to shift the beam laterally while vertical adjustment was obtained by means of the main resolving magnet M2. These magnets are shown in fig. 1 (c), which gives a general view of the whole experimental arrangement.

The energy spread of the proton beam due to ripple on the accelerating equipment, and to the use of a high-voltage ion-source, was about 30 keV. at an energy of 1 MeV. This spread could be much reduced, at the expense of intensity, by means of the horizontal slit T shown in fig. 1 (b).

§ 3. RANGE AND ENERGY CALIBRATIONS.

(a) Calibration of Range Cell.

The mean range of a group of alpha particles may be found by comparing the extrapolated numbers-range of the group with that of a standard group of alpha particles of similar range observed under the same conditions. If the two groups which are compared originate in targets or sources whose equivalent thickness is small compared with the range of the particles, it is only necessary to make a small correction for the difference in straggling of the two groups in deducing the mean range. If thick targets are used the correction from extrapolated numbers-range to mean range is that given by Livingston and Bethe (1937), if an exponential excitation function for the reaction may be assumed. If this is not so, or if the target is "medium thick", the precise correction is not easy to evaluate, but will not exceed the straggling parameter* for the group under investigation. The error in mean range introduced by the use of targets of unknown thickness will therefore not be greater than the straggling parameter, which has been taken to be 2 per cent of the mean range for alpha particles of energy 1 to 2 MeV.

In the present work the alpha-particle ranges studied were all of the order of 1 cm. of air or less, which is too far removed from those of polonium

* The straggling parameter, as defined by Livingston and Bethe (1937) and Lewis (1948) is $\sqrt{2}$ times the standard deviation of the actual ranges from the mean. The straggling coefficient Ω , used by Bøggild (1948) is equal to the standard deviation.

(3.8 cm.) and ThC (4.7 cm.) for the comparison method to be accurate. All ranges have therefore been referred to that of the group of alpha particles from the reaction ${}^9\text{Be}(p\alpha){}^6\text{Li}$. This reaction has been studied carefully by several workers, and the most reliable values for the energy releases are

2.115 ± 0.04 MeV. Allison, Skaggs, Smith (1940)

2.074 ± 0.03 MeV. Del Rosario (1948)

2.117 ± 0.01 MeV. Tollestrup, Lauritsen, Fowler (1949)

If these values are weighted (arbitrarily) in inverse proportion to the errors given, the mean value for the energy release is found to be

$$2.107 \pm 0.02 \text{ MeV.},$$

and this value has been adopted as standard. The corresponding mean alpha-particle range for a proton bombarding energy of 860 keV. and an angle of observation of 83° , is 0.93 ± 0.01 cm., according to the Cornell range-energy relation (Holloway and Livingston 1938).

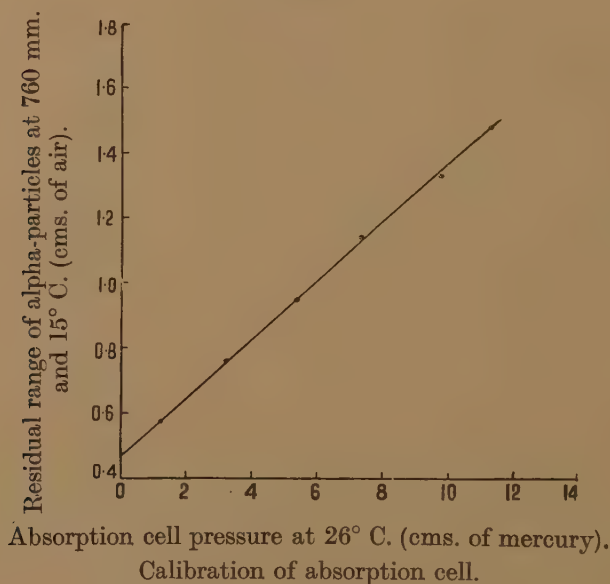
Observations were made on the alpha particles from the ${}^9\text{Be}(p\alpha)$ reaction by the method described in § 2 (c). A beryllium target of equivalent thickness less than 5 keV. for protons (as estimated by weighing) was bombarded with 860 keV. protons, and the cut-off curve, taken with the absorption cell, for the alpha particles emitted at 83° with the proton beam is that shown in fig. 5. This curve is used as a standard, and all range measurements are referred to it; the extrapolated numbers-range observed is assumed to correspond to the energy of the alpha particles emitted from the surface of the target, although owing to the nature of the excitation function for the reaction (Hornyak and Lauritsen 1948) an error of a few keV. may thus be introduced. Fig. 5 also shows a numbers-range curve for an unknown group, actually that from the ${}^{10}\text{B}(p\alpha){}^7\text{Be}$ reaction, and from the graph the difference in extrapolated numbers-range between the two groups can be read directly as a pressure difference.

The relation between this pressure difference and the actual range difference in cm. of standard air was obtained in a separate experiment with a ThC' source. The source was mounted on a micrometer screw head so that the alpha particles, after traversing a known distance in air, passed through the mica window and absorption cell to the screen. The residual range of the alpha particles as they reached the window was chosen to be about the same as that of the particles from the $(p\alpha)$ reactions under investigation, and was known to be about 0.5 cm., while the range differences were very accurately known from the micrometer readings. Several curves relating counting rate to cell pressure were obtained for different residual ranges, under the standard counting conditions, and the points where the extrapolations of these curves met the pressure axis were plotted as shown in fig. 6; the ranges were corrected to dry air at 760 mm. and 15°C . From this calibration the change of range with pressure was accurately known; the method automatically takes into account the variation in stopping power of the mica window with alpha-particle energy. By means of this calibration the differences in extrapolated

numbers-range of the alpha particles from the beryllium reaction and of those under investigation could be reduced to distances in standard air. Calibrations were made before and after each set of range measurements.

The difference between the mean range of the two groups is obtained by correcting for the difference in straggling parameters; a correction for target thickness based essentially on an estimate of this thickness in units of the straggling parameter is also made. Experimental observations on the straggling of alpha particles have recently been published by Bøggild (1948), and on the straggling of protons by Madsen and Venkateswarlu (1948). Both sets of observations suggest that 2-3 per cent of the mean

Fig. 6.



$$\frac{dR}{dp} = 0.089 \text{ cm. of air per cm. of mercury.}$$

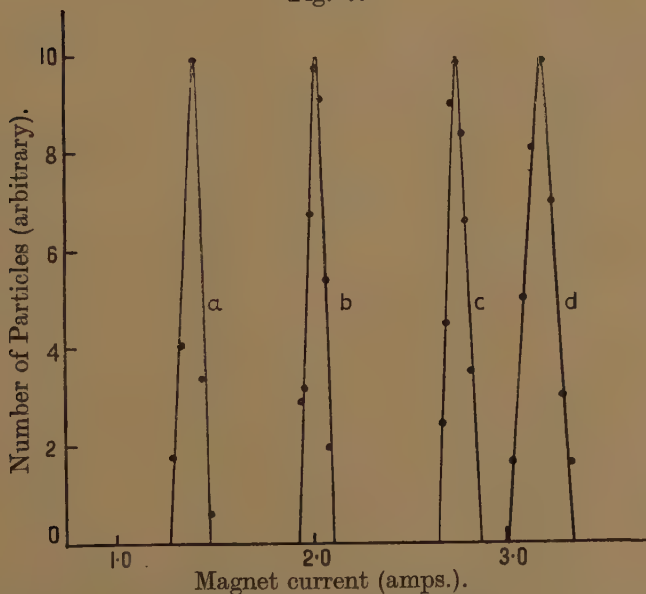
range is a safe figure to assume for the straggling parameter of alpha particles of energy 1 to 2 MeV., and this figure is in fair, though not exact, agreement with theoretical prediction (Livingston and Bethe 1937). A 2 per cent correction has been used in this work. The accuracy of range measurement by this method is estimated to be 0.02 cm. of air, or 50 keV., if the target under investigation is thin; if its thickness is unknown the accuracy is 0.03 cm., or about 70 keV. in energy. The energies corresponding to the mean ranges thus obtained are taken from the range-energy curve given by Holloway and Livingston (1948). There is some evidence that this curve gives too low a value for the energy of alpha particles of range about 1 cm. (Hacman and Haxel 1942, Bøggild 1945, Gilbert 1948, Jesse and Sadauskis 1949), but the corrections necessary are not yet well established over the whole range covered by the present experiments.

(b) Calibration of the Magnetic Analyser.

The most accurate way of measuring the energy of charged particles is deflection in an accurately known magnetic field, since this avoids dependence on a range-energy relation and on air-corrections. The analyser used in these experiments was not suitable for this purpose, owing to the effects of fringing field, but it was found possible to obtain reproducible readings of magnet current for the peaks of definite particle groups, and a calibration of the analyser in terms of magnet current was therefore made. The groups of particles used, with several bombarding energies were

- | | |
|--|---|
| (i) ${}^9\text{Be}(p\alpha){}^6\text{Li}$ | $Q=2.107$ MeV. (see § 3 (a)) |
| (ii) ${}^6\text{Li}(p\alpha){}^3\text{He}$ | $Q=3.94$ MeV. (Miller 1940, Perlow 1940) |
| (iii) ${}^9\text{Be}(pd){}^8\text{Be}$ | $Q=0.547$ MeV. (Allison, Skaggs, Smith 1940,
Del Rosario 1948, Tollestrup, Lauritsen,
Fowler 1949). |

Fig. 7.



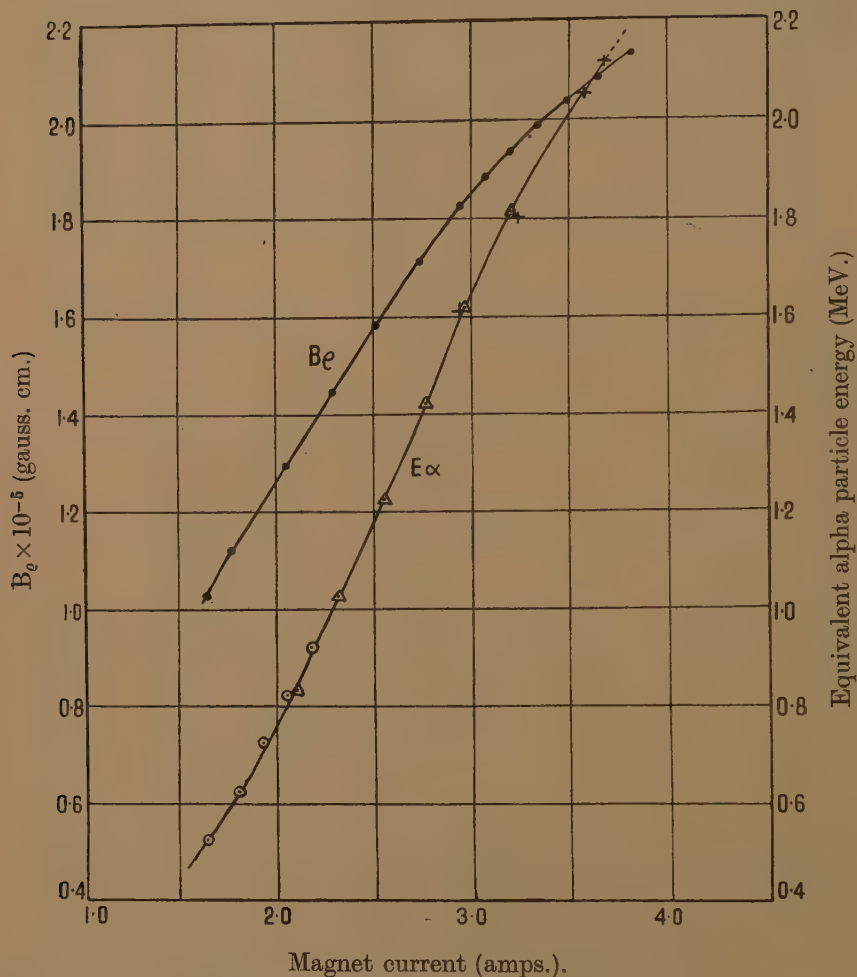
Scattered protons and deuterons.

- (a) ${}^1\text{H}$ 370 keV., (b) ${}^1\text{H}$ 800 keV., (c) ${}^2\text{H}$ 718 keV., (d) ${}^2\text{H}$ 918 keV.

It was also found possible to obtain sharp groups of scattered protons and deuterons by bombarding a thin (10 keV. equivalent) layer of gold deposited on a beryllium block in the manner described by Wilcox (1948); the bombarding voltages were calibrated by observation of the fluorine gamma-ray resonances at 873 and 667 keV. (Fowler, Lauritsen and Lauritsen 1948, Herb, Snowdon and Sala 1949) and corrections were made for target penetration (5 keV.), loss of energy in the collision (3 per cent), and difference between the atomic mass and a whole number (0.6 per cent). Fig. 7 shows a number of typical groups of scattered particles; the resolving powers calculated from these groups are listed in Table II.

Fig. 8 is the final curve, obtained by combining scattering and transmutation data, relating alpha-particle energy to magnet current. The curve has been extrapolated slightly beyond the highest experimental point by using flux-meter readings. It is believed that alpha-particle energies deduced from this curve have an accuracy of ± 50 keV., and this

Fig. 8.



Magnetic analyser calibration curve.
 $+$ ${}^9\text{Be}(p\alpha, pd)$, \triangle Scattered deuterons. \circ Scattered protons.

will determine the accuracy of the Q-values deduced from these energies if the targets used are thin. If "medium thick" targets are used, it is necessary to estimate the target thickness, and correct for it by using the excitation function for the reaction, since the peak of a group of particles corresponds to the most probable range, and is therefore much more dependent on target thickness than the extrapolated numbers-range.

For this reason rather greater reliance has been placed on range measurements in this work, except in the case of resonance reactions, where the target thickness is accurately known. Care was always taken to adjust the magnetic field according to a definite plan so as to avoid hysteresis effects; the reproducibility of the magnetic field was checked by means of a null-reading fluxmeter of the type described by Lauritsen and Lauritsen (1948). Most of the figures showing particle distributions have been plotted with magnet current as abscissa; the corresponding $B\rho$ value may be deduced from fig. 8, in which the $B\rho$ curve has been drawn to correspond to the given particle energies.

§ 4. THE REACTION $^{10}\text{B}(p\alpha)^7\text{Be}$.

The production of ^7Be in accordance with the reaction



has been reported by Roberts, Heydenburg and Locher (1938) and by Rumbaugh, Roberts and Hafstad (1938). According to the masses given by Bethe (1947) the energy release in this reaction should be 1.15 ± 0.1 MeV., and for a proton energy of 500 keV. at which the range of the scattered protons is 0.86 cm., the alpha-particle range would only be 0.55 cm. at 90° . Rumbaugh, Roberts and Hafstad attempted to observe these alpha particles, using a shallow air-filled ionization chamber as detector, from a target of ordinary boron, but failed to find any significant departure of the numbers-range distribution from the continuous distribution due to the reaction (Dee and Gilbert 1936)

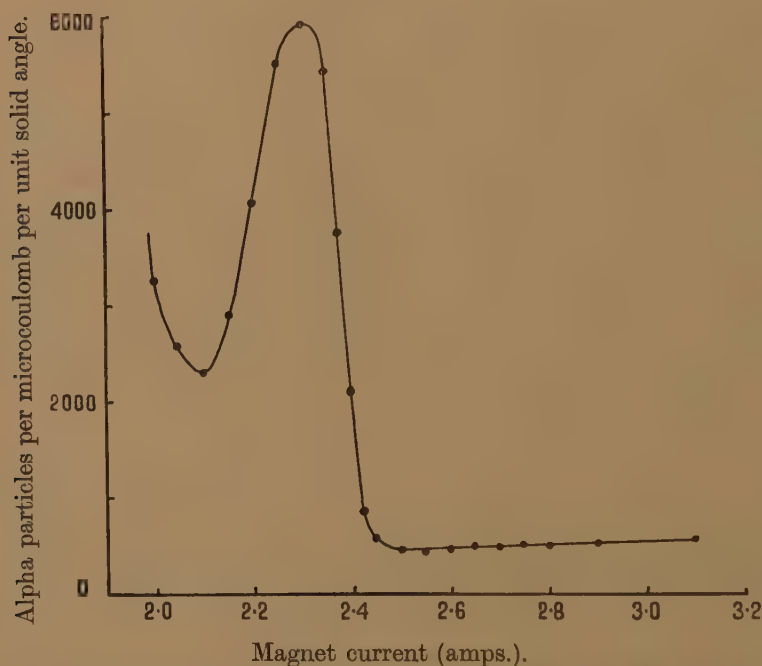


down to a bombarding energy of 300 keV., although the 43 day radio-activity of ^7Be was detected in the targets after bombardment.

In the present work, of which a preliminary account has already been published (Burcham and Freeman 1949 b), the alpha-particles from reaction (1) were observed using a target of separated ^{10}B . The magnetic analyser was used to resolve them from scattered protons, which covered a large range of momentum since the ^{10}B was deposited on a thick copper plate. The alpha particles were in the first place detected by means of the zinc sulphide screen and photo-multiplier with the screen attached directly to the analyser, as in fig. 1 (b), in order to avoid loss of range in a window. The ^{10}B sample was prepared in the small electromagnetic separator at the Atomic Energy Research Establishment, Harwell, and in the separation process a total of 2 mgm. of boron was deposited. Each trace was about $\frac{1}{4}$ in. wide and 3 ins. long and the traces were separated by $\frac{1}{2}$ in. The deposit of ^{10}B , if uniform, corresponds to an energy loss of about 30 keV. for 500 keV. protons. A target of size 17×3 mm. was cut from the ^{10}B deposit and bombarded with protons of an energy which could be varied between 250 and 1000 keV.

Fig. 9 shows the spectrum of particles observed at an angle of 83° with a beam of 630 keV. protons; the alpha-particle group is clearly resolved from the scattered protons. When the existence of this group had been established the range of the alpha particles was compared with that of the standard ${}^9\text{Be}(p\alpha)$ group by the absorption-cell method using a 1.5 mm. equivalent mica window mounted on a grid. The two numbers-range curves obtained in this comparison, which was carried out under the same geometrical conditions, but with a proton energy of 860 keV. for ${}^9\text{Be}$ and 630 keV. for ${}^{10}\text{B}$, are shown in fig. 5. From these curves and from the absorption-cell calibration shown in fig. 6, the difference between the extrapolated numbers-range of the two groups is 0.35 ± 0.01 cm. of air.

Fig. 9.



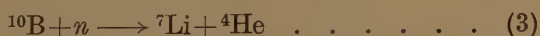
Alpha particles from the reaction ${}^{10}\text{B}(p\alpha){}^7\text{Be}$ for $E_p = 630$ keV.
Angle of observation = 83° .

The beryllium target used was known by weighing to be less than 5 keV. thick for protons, and the boron target was assumed, on the mass of the deposit, to have a thickness equal to six straggling parameters for the emitted alpha particles. A straggling correction of 0.012 cm. has been made to take account of these thicknesses and, if the mean range of the alpha particles from the ${}^9\text{Be}(p\alpha)$ reaction under the stated conditions is 0.93 ± 0.01 cm., the mean range of the ${}^{10}\text{B}(p\alpha)$ group is found to be 0.59 ± 0.02 cm. The energy of the alpha-particle group, according to the Cornell range-energy curve, is then 1.09 ± 0.03 MeV. and the energy release

in reaction (1) is found to be 1.11 ± 0.06 MeV.* This value agrees with that expected from the mass change within the limits set by the accuracy of the experiment; it differs from the value reported earlier (Burcham and Freeman 1949) owing chiefly to a revision of the energy release adopted for the ${}^9\text{Be}(p\alpha)$ reaction, which seemed desirable following the publication of the work of Tollestrup, Fowler and Lauritsen (1949). A further difference, of 0.03 MeV., arises from the introduction of the straggling correction.

The energy release was checked by the magnetic deflection method; protons scattered from a thin gold target and alpha particles from reaction (1) were observed under the same geometrical conditions, in which the beam was limited to a diameter of 5 mm. by stops. The position of the alpha-particle peak was compared directly with proton peaks for different energies, and the corresponding alpha-particle energy, for 630 keV. incident protons, was found to be 850 keV. for an angle of observation of 123° . Both these figures must be corrected for target thickness, and attempts were made to measure this directly by comparing the number of particles from the ${}^{10}\text{B}(p\alpha)$ reaction coming from the target when placed in a slow neutron flux with the number from a weighed amount of ordinary boron. The results were not accurate, owing to the difficulty of getting a thin and uniform deposit of ordinary boron, but they were consistent with a value of 30 keV. for protons (120 keV. for alpha particles), and this figure has been used. A correction of $+60$ (± 30) keV. has therefore been applied to the alpha-particle energy and -15 (± 7) keV. to the proton energy; the energy release then obtained is 1.12 ± 0.07 MeV. This agrees well with the value found from the range measurements, but is less reliable owing to the uncertainty of the target thickness correction.

Reaction (1) is in some respects analogous to the reaction



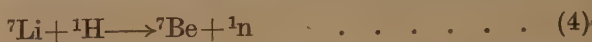
(Bøggild 1945, Gilbert 1948, Rose 1948), in which it is known that 93 per cent of the alpha particles emitted correspond to a transition to an excited state of ${}^7\text{Li}$ at 480 keV. The nuclear structure of ${}^7\text{Be}$ and ${}^7\text{Li}$ should be similar, and a search for alpha particles of both longer and shorter range than the group shown in fig. 9 was therefore made. No additional group in numbers exceeding 10 per cent of the main group could be found at higher energies (up to 2 MeV.) at any of several angles of observation in the range 48° to 134° . The search for a lower energy group was made using 230 keV. protons; fig. 10 shows the result obtained.

It can be deduced from this experiment that there is no excited state in the nucleus ${}^7\text{Be}$, which contributes to the $(p\alpha)$ reaction, between energies of 200 and 900 keV. above the ground state. The shape of the alpha-particle

* Footnote added in proof.—An accurate determination of the energy release in the ${}^{10}\text{B}(p\alpha){}^7\text{Be}$ reaction has recently been published by Chao, Lauritsen and Tollestrup (*Bull. Am. Phys. Soc.*, **24**, No. 6, p. 7, 1949). Their value is 1.148 ± 0.005 MeV.

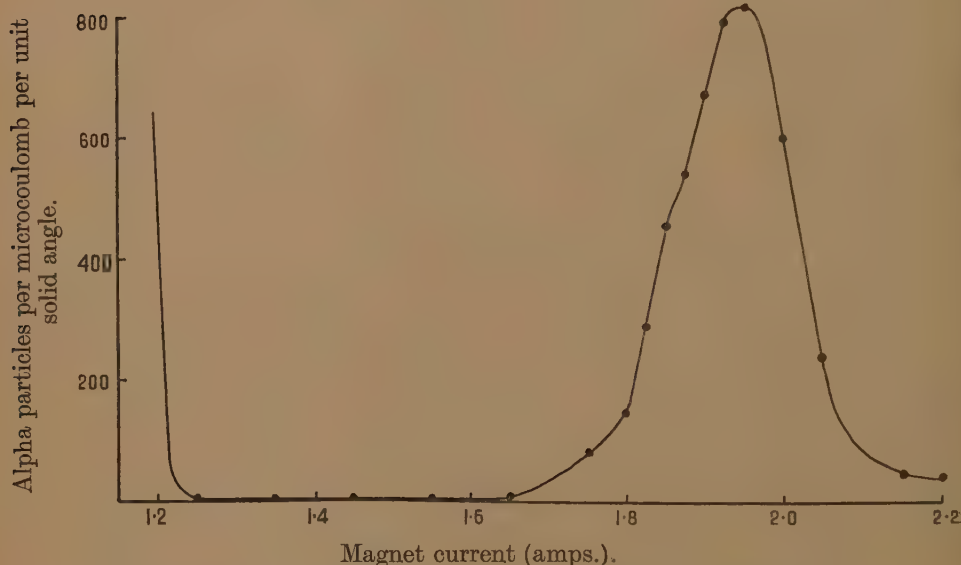
group shown in fig. 10 is not quite symmetrical, and this may mean that the group is a doublet of about 100 keV. spacing. The effect may, however, be due to irregularities of target thickness.

The experimentally observed energy releases Q_1 and Q_2 in reactions (1) and (3) should be related to the energy release Q_3 in the reaction



by the equation $Q_1 = Q_2 + Q_3$ independently of mass values, and almost independently of the alpha-particle range-energy relation. The most recent value of Q_2 is 2.78 ± 0.07 MeV. (Gilbert 1948, before correction

Fig. 10.



Alpha particles from the reaction ${}^{10}\text{B}(p\alpha){}^7\text{Be}$ for $E_p = 230$ keV.
Angle of observation = 104° .

of the range-energy relation), and Q_3 is known very accurately as $-1.65 \pm <0.01$ MeV. (Taschek and Hemmendinger 1948). The predicted value for Q_1 is then 1.13 ± 0.07 MeV., which is in agreement with that observed in the present work.

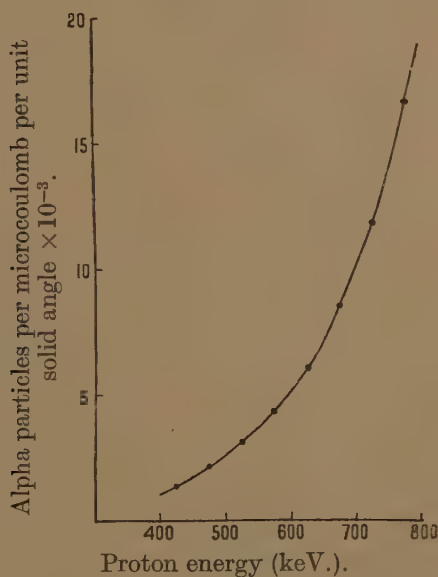
The yield of alpha particles from the ${}^{10}\text{B}$ target was investigated over a range of proton energies from 400 keV. to 800 keV. The excitation function so obtained, corrected for target thickness by the method of Tangen (1946), is shown in fig. 11.

The yield is a smoothly rising function of bombarding energy, and no levels of the compound nucleus ${}^{11}\text{C}$ can be identified in the energy range investigated. The excitation curve is of exponential form, but the accuracy of the observations, particularly at the higher energies, does not justify detailed comparison with theoretical curves.

The cross-section for the reaction was calculated from the observed excitation function and estimated target thickness, with the assumption of an isotropic angular distribution of the emitted alpha particles. At a bombarding energy of 530 keV., at which the alpha particles are well resolved from the scattered protons, the thin target yield for a solid angle of 4π is

$$Y(E) = 4.3 \times 10^{-8} \text{ alpha particles per proton.}$$

Fig. 11.



Excitation function for the reaction $^{10}\text{B}(p\alpha)^7\text{Be}$.

In calculating the cross-section for the reaction from this yield, allowance must be made for the formation of He^+ ions (which would not have been detected) as the alpha particles leave the target. From the work of Briggs (1927) it is estimated that the ratio of He^+ to He^{++} ions under the conditions of the experiment was 0.5 and the cross-section is therefore

$$\sigma(E) = 1.5 \frac{Y(E)}{N} = 1.2 \times 10^{-26} \text{ cm.}^2 \text{ for } E_p = 530 \text{ keV.}$$

where N , the number of ^{10}B nuclei per cm.^2 of the target, has been taken to be 5.4×10^{18} . The accuracy of this figure for the cross-section is chiefly dependent on the uncertainty in target thickness, and for this reason the cross-section may easily be in error by a factor of 2. The solid angle of the analyser and the target current are thought to be known to about ± 10 per cent, and the angular distribution was found to be consistent with spherical symmetry to ± 10 per cent over the angular range 48° – 134° . It is likely that most of the factors contribute to make the stated cross-section too low.

ACKNOWLEDGMENTS.

We wish to thank Sir John Cockcroft and Dr. W. D. Allen and members of his group at A.E.R.E., Harwell, who supplied the boron isotopes used in this work. We are greatly indebted to Mr. A. S. Baxter of the Cavendish Laboratory for continued help with the development of the scintillation counting technique, and to Messrs. W. Birtwhistle and D. D. Stewart for much technical assistance. One of us (J.M.F.) wishes to thank Newnham College and the Australian Council for Scientific and Industrial Research for awards.

REFERENCES.

- ALLISON, S. K., SKAGGS, L. S., and SMITH, N. M., 1940, *Phys. Rev.*, **57**, 550.
 AMALDI, E., and CACCIAPUOTI, B. N., 1947, *Phys. Rev.*, **71**, 739.
 BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9**, 69; 1947, *Elementary Nuclear Theory* (New York: John Wiley and Sons, Inc.).
 BØGGILD, J. K., 1945, *Kgl. Danske Vidensk. Selsk. Mat.-fys. Medd.*, **23**, No. 4; 1948, *Nature, Lond.*, **161**, 810.
 BRIGGS, G. H., 1927, *Proc. Roy. Soc. A*, **114**, 341.
 BROSER, I., and KALLMANN, H., 1949, *Nature, Lond.*, **163**, 20.
 BURCHAM, W. E., and DEVONS, S., 1940, *Proc. Roy. Soc. A*, **173**, 555.
 BURCHAM, W. E., and FREEMAN, J. M., 1949 a (in course of publication); 1949 b, *Nature, Lond.*, **163**, 167.
 DEE, P. I., and GILBERT, C. W., 1936, *Proc. Roy. Soc. A*, **154**, 279.
 FOWLER, W. A., LAURITSEN, C. C., and LAURITSEN, T., 1948, *Rev. Mod. Phys.*, **20**, 236.
 FREEMAN, J. M., and BAXTER, A. S., 1948, *Nature, Lond.*, **162**, 696.
 GILBERT, C. W., 1948, *Proc. Camb. Phil. Soc.*, **44**, 447.
 HACMAN, D., and HAXEL, O., 1943, *Zeits. für Phys.*, **120**, 486.
 HERB, R. G., SNOWDON, S. C., and SALA, O., 1949, *Phys. Rev.*, **75**, 246.
 HERZOG, R., 1934, *Zeits. für Phys.*, **89**, 447.
 HOLLOWAY, M. G., and LIVINGSTON, M. S., 1938, *Phys. Rev.*, **54**, 18.
 HORNYAK, W. F., and LAURITSEN, T., 1948, *Rev. Mod. Phys.*, **20**, 191.
 JESSE, W. P., and SADAUSKIS, J., 1949, *Phys. Rev.*, **75**, 1110.
 LAURITSEN, C. C., and LAURITSEN, T., 1948, *Rev. Sci. Instrum.*, **19**, 916.
 LEWIS, W. B., 1948, *Electrical Counting* (Cambridge: University Press), p. 129.
 LIVINGSTON, M. S., and BETHE, H. A., 1937, *Rev. Mod. Phys.*, **9**, 245.
 MADSEN, C. B., and VENKATESWARLU, P., 1948, *Phys. Rev.*, **74**, 1782.
 MILLER, L. C., 1940, *Phys. Rev.*, **58**, 935.
 PERLOW, G. J., 1940, *Phys. Rev.*, **58**, 218.
 DEL ROSARIO, L., 1948, *Phys. Rev.*, **74**, 304.
 ROBERTS, R. B., HEYDENBURG, N. P., and LOCHER, G. L., 1938, *Phys. Rev.*, **53**, 1016.
 ROSE, B., 1948, *Can. Jnl. Res.*, **26**, 366.
 RINGO, R., 1940, *Phys. Rev.*, **58**, 942.
 RUBIN, S., 1948, *Phys. Rev.*, **72**, 1176.
 RUMBAUGH, L. H., ROBERTS, R. B., and HAFSTAD, L. R., 1938, *Phys. Rev.*, **54**, 657.
 STEPHENS, W. E., 1934, *Phys. Rev.*, **45**, 513.
 TANGEN, R., 1946, *Kgl. Norske. Skrifter*, No. 1.
 TASCHEK, R., and HEMMENDINGER, A., 1948, *Phys. Rev.*, **74**, 373.
 TOLLESTRUP, A. V., LAURITSEN, C. C., and FOWLER, W. A., 1949, *Bull. Amer. Phys. Soc.*, **24**, 12.
 VAN ALLEN, J. A., and SMITH, N. M., 1941, *Phys. Rev.*, **59**, 501.
 WALKER, R. L., and McDANIEL, B. D., 1948, *Phys. Rev.*, **74**, 315.
 WILCOX, H. A., 1948, *Phys. Rev.*, **74**, 1743.

LXXVII. *The Characteristics of Radio-Frequency Radiation in an Ionized Gas, with Applications to the Transfer of Radiation in the Solar Atmosphere.*

By S. F. SMERD and K. C. WESTFOLD,
Radiophysics Laboratory, Council for Scientific and Industrial Research,
Commonwealth of Australia *.

[Received May 2, 1949.]

CONTENTS.

	PAGE
§1. Introduction	832
§2. The Equation of Transfer.....	833
§3. The Transfer of Radio-Frequency Radiation in the Solar Atmosphere	835
§4. The <i>Ergeibigkeit</i> of an Ionized Gas.....	836
4.1 General Considerations	836
4.2 The Effective Collision Distance	837
4.3 The Collision Frequency and Absorption Coefficient.....	839
4.4 The Frequency Distribution of Radiation emitted in an Electron-ion Encounter.....	840
4.5 The Emissivity.....	843
4.6 The <i>Ergeibigkeit</i>	843
4.7 Radiation Processes within an Ionized Gas	844
§5. Application to the Solar Atmosphere.....	848

ABSTRACT.

From consideration of the equation of transfer of radio-frequency radiation in an ionized medium it is shown that the fundamental function that determines the intensity of the radiation at any point is the *Ergeibigkeit*, the ratio of the emissivity to the product of the absorption coefficient and the square of the refractive index.

The intensity of the "quiet" radio-frequency solar radiation reaching the earth can be expressed in terms of the *Ergeibigkeit* and the optical depth of the various ray trajectories. The *Ergeibigkeit* depends on the refractive index, absorption coefficient, and emissivity of the medium. Formulæ for these are obtained in terms of the electron and ion densities and the kinetic temperature, assuming a Maxwellian distribution of velocities. The frequency distribution of radiation emitted in a classical electron-ion encounter is found for all types of encounter and an error in

* Communicated by the Council of Scientific and Industrial Research of Australia.

Kramers's formula for the emission from a parabolic encounter is discovered. The emissivity is then found by integrating over all types of encounter.

The assumption of a Maxwellian distribution implies detailed balancing of all collision processes so that the *Ergiebigkeit* should agree with that of a medium in thermodynamic equilibrium. It is shown that the formulæ for the absorption coefficient and emissivity satisfy this condition if the refractive index is unity, but in circumstances where the refractive index differs from unity a revised theory of radiation processes is required.

A heuristic theory that takes account of the effect of the surrounding particles on the absorptive and emissive processes taking place in a volume element is given. The macroscopic absorption coefficient is derived from the microscopic stimulated processes that can occur during an encounter. The coefficient of stimulated emission is related to that of spontaneous emission in the usual manner. The latter is obtained by identifying the classical spontaneous emission during an encounter with the product of the probability of the spontaneous emission of a quantum of radiation and the quantum. The absorption coefficient is then found to agree with the Lorentz formula while the emissivity has the refractive index as an additional factor.

§1. INTRODUCTION.

THE observational data on radio-frequency radiation emitted by the sun suggest that, for any one frequency, there is a level below which the mean intensity of the radiation does not fall. This is termed the thermal component of the radiation.

Enhanced intensity levels, both over long periods and in the form of short-lived variations, have to some extent been correlated with sunspot and other solar activity.

The conditions for the propagation of radio-frequency radiation in the solar atmosphere are such that propagation is severely inhibited within a level whose position ranges from chromosphere to corona with decreasing frequency. Thus, as was remarked by Martyn (1946) and independently by Ginsburg (1946), solar radio-frequency radiation observable on the earth originates in the chromosphere and corona. These may be regarded as a completely ionized hydrogen atmosphere.

In this paper we derive the characteristics of radio-frequency radiation in an ionized gas and apply the results to the problem of the transfer of the thermal radiation through the solar atmosphere. A formula for the emergent intensity of any ray is derived. It is expressed in terms of the optical depth and the *Ergiebigkeit*, which is defined as the ratio of the volume emissivity to the product of the absorption coefficient and the square of the refractive index. The refractive index comprehends the effects of scattering (cf. Darwin 1934).

We investigate the frequency-distribution of the radiation emitted during an electron-ion encounter and obtain the emissivity by integrating over all types of binary encounter, assuming the velocity distribution to be Maxwellian. The contributions from encounters between like particles are negligible. The refractive index and, in the first place, the absorption coefficient are obtained from the Lorentz theory.

§2. THE EQUATION OF TRANSFER.

The intensity distribution of radio-frequency radiation is most conveniently investigated by ray methods. The condition to be satisfied is the equation of transfer, which specifies the differential variation in intensity along a ray trajectory. The form of this equation is well known if the variation is due to energy losses from absorption and scattering by the medium and gains from its emissivity.

In our case scattering effects are taken into account by considering the variations in refractive index. Woolley (1947) has derived the equation of transfer in the appropriate form

$$\mu_f^2 \frac{d}{ds} \left(\frac{I_f}{\mu_f^2} \right) = \eta_f - \kappa_f I_f, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.1)$$

where I_f is the intensity of radiation in the frequency range $(f, f+df)$ in the direction of the ray, η_f the volume emissivity, κ_f the absorption coefficient, μ_f the refractive index of the medium and s the distance measured along a ray trajectory.

These parameters are defined such that

$$I_f df d\omega dS dt$$

is the amount of radiant energy in the range $(f, f+df)$ passing normally through an elementary area dS into an elementary solid angle $d\omega$ in time dt ,

$$\eta_f df d\omega dV dt$$

is the amount of f -radiation emitted from an elementary volume dV into $d\omega$ in time dt and

$$\kappa_f I_f df d\omega ds dS dt$$

is the amount of f -radiation passing into $d\omega$ through dS in time dt that is absorbed by the medium after traversing a distance ds along the trajectory.

We introduce the optical depth τ_f , such that

$$\tau_f(s) = \int_0^s \kappa_f ds, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.2)$$

so that equation (1.1) becomes

$$\frac{d}{d\tau_f} \left(\frac{I_f}{\mu_f^2} \right) + \frac{I_f}{\mu_f^2} = \frac{\eta_f}{\kappa_f \mu_f^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (1.3)$$

The solution is

$$\frac{I_f}{\mu_f^2} = \left(\frac{I_f}{\mu_f^2} \right)_0 \exp(-\tau_f) + \exp(-\tau_f) \int_0^{\tau_f} \frac{\eta_f}{\kappa_f \mu_f^2} \exp(\tau_f) d\tau_f, \quad (1.4)$$

giving the intensity I_f tangential to the trajectory. The subscript 0 refers to the values of I_f and μ_f where $s=0$.

In the integrand, η_f , κ_f and μ_f are associated in the form

$$F_f = \frac{\eta_f}{\kappa_f \mu_f^2}, \quad (1.5)$$

which may be called the *Ergiebigkeit* of the medium for radiation in the frequency range ($f, f+df$). It is a generalization of Schwarzschild's function to take account of variations of refractive index.

In the Lorentz theory, μ_f and κ_f are given by

$$\left. \begin{aligned} \mu_f^2 &= 1 - \frac{f_0^2}{f^2}, \\ \kappa_f &= \frac{\nu f_0^2}{c f^2 \mu_f}, \end{aligned} \right\}, \quad (1.6)$$

where

$$f_0^2 = \frac{ne^2}{\pi m}, \quad (1.7)$$

in the Gaussian system of units; ν is the number of elastic collisions per second made by an electron with the other particles of the medium, n the electron density, e and m the charge and mass of an electron and c the velocity of propagation of electromagnetic radiation in free space. The formula for μ_f^2 is valid only where both $\nu/f \ll 1$ and $f_0 < f$ (cf. Westfold, in preparation).

Under conditions of local thermodynamic equilibrium η_f is related to κ_f and μ_f by the Kirchhoff relation (v. Milne 1930)

$$\eta_f = \mu_f^2 \kappa_f B_f(T), \quad (1.8)$$

where $B_f(T)$ is the intensity of black radiation at the temperature T .

Then

$$F_f = B_f. \quad (1.9)$$

The *Ergiebigkeit* of a medium in local thermodynamic equilibrium at a temperature T is thus equal to the intensity of black radiation at that temperature.

In equilibrium the radiation temperature is identical with the kinetic temperature. For radio frequencies the Planck formula reduces to the linear Rayleigh-Jeans formula,

$$B_f = \frac{2k}{c^2} f^2 T, \quad (1.10)$$

where k is Boltzmann's constant. Then (1.14) becomes

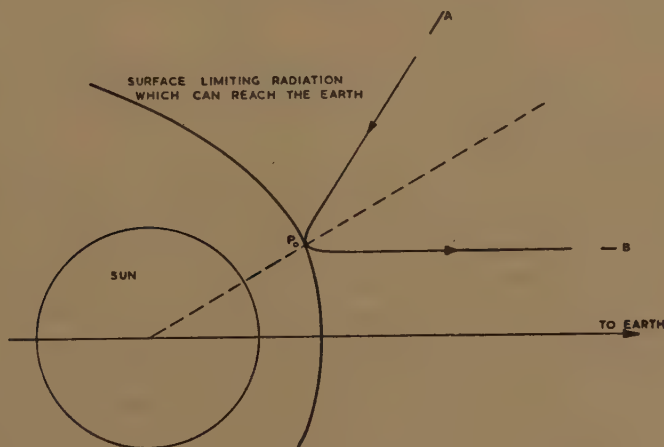
$$\frac{I_f}{\mu_f^2} = \left(\frac{I_f}{\mu_f^2} \right)_0 \exp(-\tau_f) + \frac{2k}{c^2} f^2 \exp(-\tau_f) \int_0^{\tau_f} T \exp(\tau_f) d\tau_f. \quad (1.11)$$

§3. THE TRANSFER OF RADIO-FREQUENCY RADIATION IN THE SOLAR ATMOSPHERE.

In the solar atmosphere, radio-frequency radiation of frequency f is effectively limited to the region outside the level where $f_0=f$. The refractive index is unity at the outer limits of the solar atmosphere and, apart from the effect of collisions, zero where $f_0=f$. Outside this level collisions extract part of the radiant energy traversing the medium. The effects of possible magnetic fields are neglected in this paper.

The trajectories of rays from sun to earth have been calculated by Westfold (1947), on the basis of Snell's law, extrapolating the average electron density distributions given by Baumbach (1937) and Cillié and Menzel (1935). A typical trajectory is shown in fig. 1. Radiation within the surface of revolution which is the envelope of the family of such trajectories for any frequency, cannot reach the earth. In the region outside, the formulæ (1.6) and (1.7) are valid.

Fig. 1.



The trajectory of a typical ray from sun to earth.

The intensity of a ray must be zero at the extreme point A at the effective limit of the corona, where it is directed inwards. At the other extreme point B it has the emergent value $I_f^{(e)}$. Then if s be measured from A, the first term in (1.4) vanishes and

$$I_f^{(e)} = \exp(-2\tau_f^0) \int_0^{2\tau_f^0} F_f \exp(\tau_f) d\tau_f, \quad . \quad . \quad . \quad (2.1)$$

where τ_f^0 is the optical depth from an extreme point to the innermost point P_0 attained by a trajectory, often wrongly termed the "point of reflection". The relation (2.1) is the only proper basis for the interpretation of terrestrial radio-frequency observations in terms of a ray theory.

According to the present physical evidence it is a good approximation to regard the solar atmosphere—chromosphere and corona—as completely ionized hydrogen. The electron kinetic temperature ranges from 10^{40}K. in part of the chromosphere to about 10^{60}K. in the inner corona.

With the assumption of local thermodynamic equilibrium (2.1) becomes

$$I_f^{(e)} = \frac{2k}{c^2} f^2 \exp(-2\tau_f^0) \int_0^{2\tau_f^0} T \exp(\tau_f) d\tau_f. \quad (2.2)$$

The validity of this assumption and hence the reduction of (2.1) to (2.2) depend on the extent to which radiation outside the level $f_0=f$ may be regarded as enclosed, or how nearly it resembles equilibrium radiation. The chief contribution to the emissivity will be from accelerating electrons in free-free electron-proton encounters. In sub-section 4.4 we investigate the frequency distribution of radiation emitted in such an encounter using classical electromagnetic theory. For radio frequencies the results should, by the correspondence principle, be equivalent to those of quantum theory. The final values of emissivity, absorption coefficient and *Ergiebigkeit* assuming a Maxwellian distribution of velocities are obtained in sub-section 4.7.

§4. THE *ERGIEBIGKEIT* OF A IONIZED GAS.

4.1. General Considerations.

The *Ergiebigkeit* of an ionized gas is determinate at all points if the distributions of μ_f , κ_f and η_f are known. From (1.6) and (1.7) we see that μ_f depends only on the electron density while κ_f depends on μ_f and ν . Both ν and η_f must depend on the nature of the individual collision processes. We shall suppose that the gas is binary, consisting only of electrons of mass m and charge e and positive ions of mass M and charge Ze .

Now the dynamics of two interacting molecules of masses m and M are determined by G , the velocity of their centre of mass (which remains constant), g the initial velocity of m relative to M , b the perpendicular distance between M and either asymptote of the trajectory of m relative to M , and the law of force between the two molecules. If $M \gg m$ it can be shown (*cf.* Chapman and Cowling 1939 a) that the number of binary collisions per unit volume and time such that b , g and G lie in the ranges $(b, b+db)$, $(g, g+dg)$ and $(G, G+dG)$ is

$$nN \frac{(mM)^{\frac{3}{2}}}{(2\pi kT)^3} \exp[-(mg^2 + MG^2)/2kT] 2\pi gb db \cdot 4\pi g^2 dg \cdot 4\pi G^2 dG, \quad (3.1)$$

assuming a Maxwellian distribution of the velocities. Here n and N are the number densities of molecules of masses m and M respectively and T is the kinetic temperature of the gas. Both G and g range from zero to infinity and b ranges from zero to the effective collision distance σ .

For rigid elastic spherical molecules σ is the mean diameter of the two molecules; if $b > \sigma$ no collision can occur. For molecules that are centres of force, the mean distance between neighbouring molecules of the medium

must be such that the mutual influence of two interacting molecules is negligible before they are affected by the fields of other molecules, if the proportion of multiple to binary encounters is to be small. Those interaction forces of kinetic theory that vary as higher than the second inverse power of the distance between two molecules decrease so rapidly that σ may be taken as infinity. Electrostatic forces, however, decrease so slowly with distance that the mutual influence of two interacting molecules is still appreciable at separations where they have also come under the influence of other molecules. Chapman and Cowling (1939 b) show that it is still possible to derive approximate expressions for the coefficients of viscosity, conduction and diffusion using (3.1). The condition is that encounters involving deflections greater than some prescribed, small value should occur for values of b which are small compared with the mean distance between neighbouring molecules. If we compare the expression for the coefficient of mutual diffusion of the two gases with the corresponding expression for rigid spherical molecules we can infer an effective collision distance σ for molecules that are centres of electrostatic force.

We can then write down the expression for ν , from (3.1),

$$\nu = 32N \frac{(mM)^{\frac{3}{2}}}{(2kT)^3} \int_0^\infty \int_0^\infty \int_0^\sigma G^2 g^3 b \exp[-(mg^2 + MG^2)/2kT] db dg dG. \quad (3.2)$$

Again, if $Q_f df$ is the energy radiated in the frequency range $(f, f+df)$ during a collision, the spontaneous emission per unit volume and time is given by

$$4\pi n_f df = 32 nN \frac{(mM)^{\frac{3}{2}}}{(2kT)^3} \int_0^\infty \int_0^\infty \int_0^\sigma Q_f df G^2 g^3 b \exp[-(mg^2 + MG^2)/2kT] db dg dG. \quad (3.3)$$

Although the mean collision distance σ is the maximum value of b in all cases where rigid elastic spherical molecules are involved, we shall see in the next sub-section that it is not possible to regard this inferred effective collision distance σ as an equivalent concept. It will then be necessary to alter the terminal σ in (3.3).

In both (3.2) and (3.3) we have assumed that the energy radiated is a very small fraction of the total energy of an interacting pair so that b , g and G remain sensibly unaltered throughout an encounter. Again, (3.3) has been written down on the assumption that each electron-ion pair makes its contribution to the emissivity independently of the presence of the surrounding particles of the medium. In sub-section 4.7 we shall adduce reasons why a factor that takes account of their presence under circumstances where the refractive index differs from unity should be introduced.

4.2. The Effective Collision Distance.

As outlined in the previous sub-section, an effective collision distance in an ionized medium can be inferred by considering a transport phenomenon which involves the constituent ions and electrons and comparing

the result with that for a binary gas consisting of rigid elastic spherical molecules. Such a phenomenon is diffusion. If D is the coefficient of mutual diffusion for a mixture of two gases, first approximations are

$$D = \frac{3}{16(n+N)\sigma^2} \left\{ \frac{2kT(m+M)}{\pi m M} \right\}^{\frac{1}{2}} \quad \dots \quad (3.4)$$

for rigid spherical molecules (Chapman and Cowling 1939 c) and

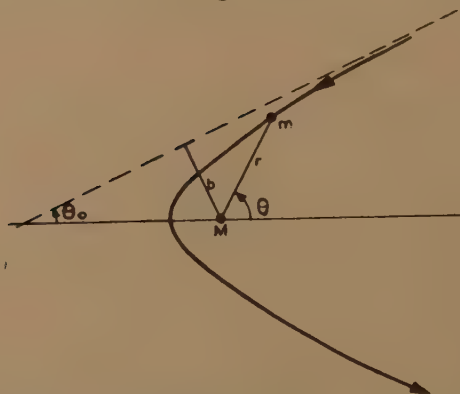
$$D = \frac{3}{16(n+N)} \left\{ \frac{2kT(m+M)}{\pi m M} \right\}^{\frac{1}{2}} \left(\frac{2kT}{Ze^2} \right)^2 / A_1(2) \quad \dots \quad (3.5)$$

for an ionized gas (Chapman and Cowling 1939 d). If v_{01} is the maximum value of

$$v_0 = \frac{mg^2b}{Ze^2}, \quad \dots \quad (3.6)$$

$$A_1(2) = \ln(1 + v_{01}^2). \quad \dots \quad (3.7)$$

Fig. 2.



The trajectory of the electron relative to the positive ion.

In (3.5) mg^2 in $A_1(2)$ is replaced by its mean value $4kT$.

Comparing (3.4) and (4.5) we find that the effective collision distance for the ionized gas is given by

$$\sigma = \{A_1(2)\}^{\frac{1}{2}} \frac{Ze^2}{2kT} \quad \dots \quad (3.8)$$

When we consider the dynamics of an encounter between an electron and a positive ion in sub-section 4.4, we shall see that if θ_0 is the angle between the initial asymptote of the trajectory of the electron relative to the ion and the apse-line (fig. 2),

$$v_0 = \tan \theta_0. \quad \dots \quad (3.9)$$

Before carrying out the integration with respect to g to get (3.5) Chapman and Cowling argue that since $A_1(2)$ is a slowly varying function of v_{01} , it is permissible to replace b_{\max} by d , the mean distance between

4.4. The Frequency Distribution of Radiation emitted in an Electron-ion Encounter.

According to classical electromagnetic theory, the instantaneous rate at which radiation is emitted by an accelerating charge is

$$\frac{dQ}{dt} = \frac{2e^2}{3c^3} j^2, \quad \dots \dots \dots (3.13)$$

where j is the instantaneous acceleration and e the charge in Gaussian units. In an electron-ion encounter both particles are accelerated in the Coulomb field of electrostatic attraction. Since their accelerations at any instant are inversely proportional to their masses m and M , the effect of the accelerating positive ion can be neglected in comparison with that of the electron. It is for this reason also that the positive ions of the medium make no appreciable contribution to the formulæ (1.6).

The frequency distribution of radiation from an encounter was investigated by Kramers (1923) who made a Fourier analysis of the total radiation Q emitted by the electron, using classical electromagnetic theory.

Taking polar coordinates, r, θ relative to the ion, the electron describes the conic

$$\frac{l}{r} = 1 - \epsilon \cos \theta,$$

where

$$l = b\nu_0$$

and

$$\epsilon^2 = 1 + \nu_0^2. \quad \dots \dots \dots (3.14)$$

The parameters g, b and ν_0 which determine the semi-latus rectum l and eccentricity ϵ , have already been defined. The angle θ is measured from the apse-line (fig. 2). Since $\epsilon > 1$, the conic is a hyperbola. The initial value θ_0 of θ is $\arccos 1/\epsilon$, which by (3.14) is also $\arctan \nu_0$, proving the relation (3.9). The final value of θ is $2\pi - \theta_0$ whence the total deflection of the electron is $\pi - 2\theta_0$. If g and b are such that $\nu_0 \ll 1$, $\epsilon \simeq 1$ and the orbit is nearly parabolic. The deflection is then nearly π . If g and b are such that $\nu_0 \gg 1$ the deflection is very small. These are the extreme cases which can occur.

The position of m relative to M at any instant is found by integrating the angular momentum relation

$$r^2 \dot{\theta} = gb. \quad \dots \dots \dots (3.15)$$

Taking $t=0$ at the apse $\theta=\pi$, the result is

$$\frac{gt}{b} = \frac{2\epsilon \tan \frac{1}{2}\theta}{(\epsilon-1) + (\epsilon+1)\tan^2 \frac{1}{2}\theta} + \frac{1}{\sqrt{(\epsilon^2-1)}} \ln \left| \frac{\sqrt{(\epsilon-1)} + \sqrt{(\epsilon+1)} \tan \frac{1}{2}\theta}{\sqrt{(\epsilon-1)} - \sqrt{(\epsilon+1)} \tan \frac{1}{2}\theta} \right|. \quad \dots \dots (3.16)$$

The duration of an encounter is then from $t=-\infty$ to $t=+\infty$ and the total radiation emitted is

$$Q = \frac{2e^2}{3c^3} \int_{-\infty}^{\infty} j^2 dt. \quad . \quad . \quad . \quad . \quad . \quad (3.17)$$

Resolving j into components $j_{||}$ and j_{\perp} parallel and perpendicular to the apse line we have

$$\left. \begin{aligned} j_{||} &= -\frac{Ze^2}{mr^2} \cos \theta, \\ j_{\perp} &= -\frac{Ze^2}{mr^2} \sin \theta, \\ j^2 &= j_{||}^2 + j_{\perp}^2. \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad . \quad (3.18)$$

and

Now if we analyse $j_{||}(t)$ and $j_{\perp}(t)$ into Fourier components of amplitudes $J_{||}(f)$ and $J_{\perp}(f)$ such that,

$$\left. \begin{aligned} j_{||} &= \int_{-\infty}^{\infty} J_{||} \exp(-2\pi i f t) df, \\ j_{\perp} &= \int_{-\infty}^{\infty} J_{\perp} \exp(-2\pi i f t) df, \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad . \quad (3.19)$$

we have the inverse relations

$$\left. \begin{aligned} J_{||} &= \int_{-\infty}^{\infty} j_{||} \exp(2\pi i f t) dt, \\ J_{\perp} &= \int_{-\infty}^{\infty} j_{\perp} \exp(2\pi i f t) dt. \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad . \quad (3.20)$$

Then by the Parseval formula for Fourier integrals,

$$\int_{-\infty}^{\infty} j^2 dt = 2 \int_0^{\infty} (|J_{||}|^2 + |J_{\perp}|^2) df. \quad . \quad . \quad . \quad . \quad (3.21)$$

By the definition of Q_f given at the end of sub-section 4.1,

$$Q = \int_0^{\infty} Q_f df. \quad . \quad . \quad . \quad . \quad . \quad (3.22)$$

Hence we may write

$$Q_f = \frac{4e^2}{3c^3} (|J_{||}|^2 + |J_{\perp}|^2). \quad . \quad . \quad . \quad . \quad . \quad (3.23)$$

Using (3.15) we can transform the variable of integration in (3.20) from t to θ to find

$$\left. \begin{aligned} J_{||} &= -\frac{Ze^2}{mgb} \int_{\theta_0}^{2\pi-\theta_0} \exp(2\pi i f t) \cos \theta d\theta, \\ J_{\perp} &= -\frac{Ze^2}{mgb} \int_{\theta_0}^{2\pi-\theta_0} \exp(2\pi i f t) \sin \theta d\theta, \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad . \quad (3.24)$$

where t is a function of θ given by (3.16).

These integrals cannot be expressed exactly in terms of known functions. Kramers considered the two extreme cases mentioned above. In the near-parabolic case he took $\theta_0=0$ in the terminals of (3.24) and obtained results in terms of Hankel functions of order $\frac{1}{2}$. The corresponding formula for Q_f forms the basis of his well-known investigation of the effective cross-section for capture of the electron into one of the Bohr quantum orbits.

He obtained an improved formula for the cross-section by investigating Q_f also for the case of zero deflection. He then took this as typical over the range $\frac{1}{4}\pi < \theta_0 < \frac{1}{2}\pi$ and the parabolic formula as typical over the range $0 < \theta_0 < \frac{1}{4}\pi$. For zero deflections $J_{||}$ and J_{\perp} were expressed in terms of Hankel functions of order zero and unity.

Now, for conditions such as those in the solar atmosphere, the arguments of these functions are so small that they may be approximated by the first terms of their power series expansions. Kramers's formulæ then give $Q_f \propto f^{\frac{2}{3}}$ and Q_f independent of f , for the parabolic and zero deflection cases respectively. This procedure, however, is equivalent to taking the first non-zero terms in the series resulting from substitution of the power series expansions of $\exp(2\pi i f t)$ in the integrands of (3.24). Under such conditions we should therefore be able to derive an approximate expression for Q_f directly from (3.23) and (3.24), valid not only in the extreme cases considered by Kramers, but throughout the range $0 < \theta_0 < \frac{1}{2}\pi$. Physically, it is apparent that the extreme portions of the trajectory are the least important as far as radiation is concerned. An examination of the right side of (3.16) shows that gt/b is not greater than, say, 10 over that part of the trajectory where most of the deviation takes place. If $fb/g \ll 1$, we can therefore replace $\exp(2\pi i f t)$ by 1 in the integrals (3.24). Replacing g by its mean value and b by d , the condition may be written

$$f \ll 10^6 n^{\frac{1}{2}} T^{\frac{1}{2}}. \quad \dots \quad (3.25)$$

Then we get the approximate result

$$\left. \begin{aligned} J_{||} &= \frac{2Ze^2}{mgb} \sin \theta_0, \\ J_{\perp} &= 0, \end{aligned} \right\} \quad \dots \quad (3.26)$$

where $\sin \theta_0$ can be expressed in terms of g and b by means of (3.6) and (3.9). The corresponding expression for Q_f is then

$$Q_f = \frac{16e^2 g^2}{3c^3} \cdot \frac{1}{1 + (mg^2 b / Ze^2)^2} \cdot \dots \quad (3.27)$$

The chief characteristic of this formula is its independence of the frequency f . When the deflection is small, ν_0 is large and (3.27) becomes, approximately,

$$Q_f = \frac{16Z^2 e^6}{3c^3 m^2 g^2 b^2}, \quad \dots \quad (3.28)$$

agreeing with Kramers's result for zero deflection. When the trajectory is nearly parabolic, ν_0 is small and approximately

$$Q_f = \frac{16e^2g^2}{3c^3} \quad (3.29)$$

This does not agree with Kramers's frequency-dependent result. The reason can be traced to Kramers's method of approximation. In the limits of the integrals corresponding to (3.24), he took $\theta_0=0$ when it should have been retained as a small quantity, to the first order equal to ν_0 . His procedure is equivalent to the invalid procedure of putting $\sin \theta_0=0$ in (3.26) for the parabolic case. Then the next approximation would involve taking the next term in the expansion of $\exp(2\pi if t)$ and would give the frequency-dependent result $Q_f \propto f^{\frac{1}{2}}$. The consequences of Kramers's error will be considered in the following paper.

4.5. The Emissivity.

Substituting (3.27) into the integral (3.3), we find on integrating with respect to G and b ,

$$\eta_f = \frac{16nNZ^2e^6}{3\pi^{\frac{1}{2}}c^3m^2} \left(\frac{m}{2kT} \right)^{\frac{3}{2}} \int_0^\infty A_1(2)g \exp(-mg^2/2kT) dg, \quad (3.30)$$

where $A_1(2)$ is the same function as is defined by (3.7). Again, to get a result in terms of known functions, we make use of the fact that $A_1(2)$ is a slowly varying function of ν_{01} and replace mg^2 in ν_{01} by its mean value $4kT$. We then find

$$\eta_f = \frac{8nNZ^2e^6A_1(2)}{3c^3(2\pi m^3kT)^{\frac{1}{2}}} \quad (3.31)$$

We choose the minimum deflection in the same way as we did for diffusion, so that $A_1(2)$ in (3.31) is given by (3.7) and (3.10). Thus η_f depends on the electron and ion densities and the kinetic temperature of the gas. We have already noted that an additional factor representing the effect of the surrounding medium is required in (3.31).

4.6. The *Ergiebigkeit*.

An expression for the *Ergiebigkeit* of an ionized gas can now be obtained by substituting (3.12) and (3.31) into (1.5). The result is

$$F_f = \frac{8k}{3c^2\mu_f} f^2 T, \quad (3.32)$$

where μ_f is a function of n given by (1.6). It should be noted that this formula is independent of the value of $A_1(2)$. It depends only on the electron density and the kinetic temperature.

The ratio of the *Ergiebigkeit* of the gas to that of matter in local thermodynamic equilibrium at the same temperature is obtained by comparison with (1.10). We find

$$F_f = \frac{4}{3\mu_f} B_f \quad (3.33)$$

The presence of the factor μ_f in the denominator of (3.33) excites some suspicion. If we imagine our medium to be of great extent, at some central point the conditions should be similar to those within an enclosure at a temperature T . Here we should have $F_f = B_f$ and this cannot be obtained from (3.33). Indeed, the assumption of a Maxwellian distribution of velocities implies detailed balancing which must lead to the equilibrium result. The numerical factor $4/3$ is of little account. It can be ascribed to the free-path methods inherent in the Lorentz investigation of the absorption coefficient κ_f .

In the next sub-section we shall investigate the theory of free-free radiation transition processes in a manner analogous to that of Einstein (1917). Certain heuristic modifications of the analysis will be found necessary in order to take account of a refractive index different from unity. Such a situation has apparently not hitherto been considered.

4.7. Radiation Processes within an Ionized Gas.

Here we consider the microscopic absorptive and emissive processes which take place within an ionized gas. Since all the electrons are free, the primary process is that of an electron-ion encounter.

The variables specifying an encounter are g and b . By integrating (3.1) over the range of G we find that the number of (g, b) encounters per unit volume and time is

$$\phi(g, b) db dg = 8\pi^{\frac{1}{2}} n N \left(\frac{m}{2kT} \right)^{\frac{3}{2}} b g^3 \exp(-mg^2/2kT) db dg. \quad (3.34)$$

Now we suppose that during an encounter the mutual energy of an interacting pair changes from $\frac{1}{2}mg_1^2$ to $\frac{1}{2}mg_2^2$ with the absorption of a quantum hf from the radiation field. Thus

$$hf = \frac{1}{2}m(g_2^2 - g_1^2), \quad \dots \dots \dots (3.35)$$

where h is Planck's constant. The probability of this transition $1 \rightarrow 2$ occurring during an encounter is assumed to be proportional to the intensity I_f of f -radiation; it may be written

$$\beta_{12} I_f. \quad \dots \dots \dots (3.36)$$

The converse transition $2 \rightarrow 1$ with the emission of hf may occur spontaneously with the probability

$$\alpha_{21}, \quad \dots \dots \dots (3.37)$$

or it may be stimulated by the radiation field and have the probability

$$\beta_{21} I_f. \quad \dots \dots \dots (3.38)$$

The probability coefficients β_{12} , α_{21} and β_{21} are functions of the encounter variables g, b in the two states.

Now, under equilibrium conditions, there must be a detailed balancing of the number of processes $1 \rightarrow 2$ against the number of processes $2 \rightarrow 1$ per unit volume and time, and so on. If we assume that each process takes place *independently of the presence of the surrounding particles of the medium* we

must have

$$\phi_1 db_1 dg_1 \beta_{12} I_f = \phi_2 db_2 dg_2 (\alpha_{21} + \beta_{21} I_f).$$

Then, since from (3.34) and (3.35)

$$\left. \begin{aligned} \frac{\phi_1 db_1 dg_1}{\phi_2 db_2 dg_2} &= \frac{b_1 db_1 g_1^3 dg_1}{b_2 db_2 g_2^3 dg_2} \exp(hf/kT), \\ I_f &= \frac{\alpha_{21}/\beta_{21}}{\frac{b_1 db_1 g_1^3 dg_1}{b_2 db_2 g_2^3 dg_2} \frac{\beta_{12}}{\beta_{21}} \exp(hf/kT) - 1} \end{aligned} \right\} \quad (3.39)$$

We now introduce Einstein's plausible "boundary condition" that when $T \rightarrow \infty$, $I_f \rightarrow \infty$. This yields the result

$$\frac{b_1 db_1 g_1^3 dg_1}{b_2 db_2 g_2^3 dg_2} \beta_{12} = \beta_{21}, \quad (3.40)$$

whence

$$I_f = \frac{\alpha_{21}/\beta_{21}}{\exp(hf/kT) - 1}.$$

The simplest procedure now is to appeal to the correspondence principle. To agree with the Rayleigh-Jeans formula (1.10) for I_f , valid when $hf \ll kT$, we must have

$$\alpha_{21}/\beta_{21} = 2hf^3/c^2. \quad (3.41)$$

Thus we arrive at the result that

$$\begin{aligned} I_f &= \frac{2hf^3/c^2}{\exp(hf/kT) - 1} \\ &= B_f, \end{aligned} \quad (3.42)$$

the Planck's intensity function.

We can find the absolute values of the probability coefficients for the type of encounter under consideration by identifying the spontaneous emission $Q_f df$ found classically in sub-section 4.4 with the product of the probability of the spontaneous emission of a quantum during an encounter and the quantum. Thus

$$\alpha_{21} = Q_f df/hf \quad (3.43)$$

and β_{21} and β_{12} are determined by (3.41) and (3.40).

We can now determine the macroscopic absorption coefficient κ_f . As Martyn (1948) has pointed out, it represents the net effect stimulated by the radiation field, expressed as an absorption. The net loss of the f -radiation incident normally on a surface dS within a solid angle $d\omega$ in time dt after travelling a distance ds is

$$\left\{ \int_0^\infty \int_0^{\max} \phi_1 \beta_{12} db_1 dg_1 - \int_0^\infty \int_0^{\max} \phi_2 b_{21} db_2 dg_2 \right\} I_f hf (d\omega/4\pi) ds dS dt.$$

In §1 we saw that, in terms of κ_f , this is given by

$$\kappa_f I_f df d\omega ds dS dt.$$

Hence, using (3.39) and (3.40), we get

$$4\pi\kappa_f df = hf (\exp(hf/kT) - 1) \int_0^\infty \int_0^{b_{\max}} \phi_2 \beta_{21} db_2 dg_2. \quad (3.44)$$

The integral with respect to b and g corresponds with the sum $\sum n_2 B_{21}$ over all states 2 from which a quantum hf is emitted, in the usual Einstein treatment of transitions between stationary states*.

Now, substituting our formula (3.27) for Q_f , we get

$$\left. \begin{aligned} \alpha_{21} &= \frac{16e^2 g_2^2}{3c^3 hf} \frac{df}{1 + \nu_{02}^2}, \\ \beta_{21} &= \frac{8e^2 g_2^2}{3ch^2 f^4} \frac{df}{1 + \nu_{02}^2} \end{aligned} \right\} \dots \dots \dots (3.45)$$

Then from (3.44), (3.34) and (3.45), after integrating with respect to b_2 , we get

$$\kappa = \frac{8nNZ^2e^6}{3\pi^{\frac{1}{2}}cm^2hf^3} \left(\frac{m}{2kT}\right)^{\frac{3}{2}} (\exp(hf/kT) - 1) \int_0^\infty A_1(2)g \exp(-mg^2/2kT) dg. \quad (3.46)$$

We resolve the remaining integral in the same way as in sub-section 4.5 and, since $hf \ll kT$, replace $[\exp(hf/kT) - 1]$ by its first approximation hf/kT .

The final result is

$$\kappa_f = \frac{4nNZ^2e^6 A_1(2)}{3\{2\pi(mkT)^{\frac{3}{2}}\}^{\frac{1}{2}}cf^2}, \quad (3.47)$$

where $A_1(2)$ is given by (3.7) and (3.10).

The results (3.42) and (3.47) are independent of the refractive index of the medium. They are, in fact, correct only if $\mu_f = 1$. If the refractive index differs from unity, (3.42) should be replaced by

$$I_f = \mu_f^2 B_f \quad (3.48)$$

and there should be a factor μ_f in the denominator of (3.47). The former correction can be seen from (1.3), (1.5) and (1.9) for a medium in thermodynamic equilibrium, where I_f/μ_f^2 must remain constant. The latter is obtained from comparison with (3.12). The additional factor $4/3$, which we have been missing in (3.33), has come from (3.45), (3.41) and the identification of a $\alpha_{21}hf$ with $Q_f df$ which was derived independently of free-path methods.

The only place where the foregoing argument can be modified to obtain the desired result is in the assumption that the individual processes take place independently of the presence of the surrounding particles of the medium. It is reasonable, however, to postulate that the number of transitions, spontaneous or stimulated, that take place per unit volume and time should depend not only on the probability of a transition in an isolated encounter and the number of encounters per unit volume and time, but

* $B_{21}I_f dt$ is the probability of a stimulated transition $2 \rightarrow 1$ occurring within time dt and n_2 is the number of atoms per unit volume in state 2.

also on a macroscopic parameter that specifies some inter-relation between the particles of the medium. This parameter should have a negligible effect when the particle density is so low or the frequency so high that the velocity of propagation of the radiation is not sensibly affected by the medium. From purely heuristic considerations we choose the refractive index μ_f to fulfil this rôle in both stimulated and spontaneous transitions.

To correct the formula (3.47) for κ_f we postulate that the number of absorptions $1 \rightarrow 2$ per unit volume and time is

$$\phi_1 db_1 dg_1 \beta_{12} I_f / \mu_f \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.49)$$

and that the number of stimulated emissions $2 \rightarrow 1$ per unit volume and time is

$$\phi_2 db_2 dg_2 \beta_{21} I_f / \mu_f \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.50)$$

Then instead of (3.47) we obtain

$$\kappa_f = \frac{4nNZ^2e^6A_1(2)}{3\{2\pi(mkT)^3\}^{\frac{1}{2}}cf^2\mu_f} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.51)$$

Again, to correct (3.42) to (3.48), we postulate that the number of spontaneous emissions $2 \rightarrow 1$ per unit volume and time is

$$\phi_2 db_2 dg_2 \alpha_{21} \mu_f \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.52)$$

Now in sub-section 4.5 the emissivity η_f was derived on the assumption that the emission from each encounter was contributed independently to η_f . We see now that (3.31) requires the additional factor μ_f to give

$$\eta_f = \frac{8nNZ^2e^6A_1(2)\mu_f}{3c^3(2\pi m^3kT)^{\frac{1}{2}}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.53)$$

The formulæ (3.48), (3.51) and (3.53) now supersede (3.42), (3.12) and (3.31).

The energy density ρ_f is related to I_f by the equation

$$\rho_f = \frac{4\pi}{u_f} I_f \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.54)$$

where u_f is the velocity of energy propagation of f -radiation. In an ionized gas

$$u_f = c\mu_f \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.55)$$

whence

$$\rho_f = \frac{4\pi\mu_f}{c} B_f \quad . \quad . \quad . \quad . \quad . \quad . \quad (3.56)$$

This relation is different from that usually quoted for energy density in a medium possessing a refractive index. There the velocity of energy propagation is taken to be the same as the phase velocity, so that

$$u_f = c/\mu_f,$$

whence

$$\rho_f = \frac{4\pi\mu_f^3}{c} B_f.$$

Now, for the *Ergiebigkeit*, we get the result

$$\begin{aligned} F_f &= \frac{2k}{c^2} f^2 T, \\ &= B_f. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \end{aligned} \tag{3.57}$$

where $hf \ll kT$. Thus we have verified that the radio-frequency *Ergiebigkeit* of an ionized gas whose distribution of velocities is approximately Maxwellian is equal to that of a medium in thermodynamic equilibrium.

§5. APPLICATION TO THE SOLAR ATMOSPHERE.

To obtain expressions for κ_j and η_j applicable to our model of the solar atmosphere we need only put $n=N$ and $Z=1$ in (3.51) and (3.53). The expression for μ_j is already given by (1.6). Thus all the characteristics of radio-frequency radiation in the solar atmosphere are determined by the distributions of n and T . Finally, the emergent intensity of any ray is, by (2.1) and (3.57),

$$I_f^{(e)} = \exp(-2\tau_f^0) \int_0^{2\tau_f^0} B_f \exp(\tau_f) d\tau_f, \quad . \quad . \quad (4.1)$$

which is equivalent to (2.2).

An investigation of the distribution of emergent intensities, from centre to limb of the solar disk, of sun-earth rays of various radio frequencies will shortly be published.

ACKNOWLEDGMENT.

The work described in this paper is part of the research programme of the Division of Radiophysics of the Council for Scientific and Industrial Research, Australia. Valuable criticism has been received from Mr. R. G. Giovanelli of the Division of Physics.

REFERENCES.

- BAUMBACH, S., 1937, *Astr. Nachr.*, **263**, 121.
CHAPMAN, S., and COWLING, T. G., 1939 a, *The mathematical theory of non-uniform gases* (Cambridge: University Press), Section 5.2; 1939 b, *Ibid.*, Section 10.33; 1939 c, *Ibid.*, Section 10.22; 1939 d, *Ibid.*, Section 10.33.
CILLIÉ, G. G., and MENZEL, D. H., 1935, *Harvard College Observatory Circular* 410.
DARWIN, C. G., 1924, *Trans. Camb. Phil. Soc.*, **23**, 137.
EINSTEIN, A., 1917, *Phys. Zeits.* **18**, 121.
GINSBURG, V. L., 1946, *C.R. Acad. Sci., U.S.S.R.*, **52**, 487.
KRAMERS, H. A., 1923, *Phil. Mag.*, **46**, 836.
MARTYN, D. F., 1946, *Nature, Lond.*, **158**, 632; 1948, *Proc. Roy. Soc. A*, **193**, 44.
MILNE, E. A., 1930, *Handbuch d. Astrophysik*, **3**, 80.
WESTFOLD, K. C., 1947, *C.S.I.R. Radiophysics Laboratory*, RPL. 7; 1949, *The interpretation of magneto-ionic theory* (in preparation).
WOOLLEY, R. V. D. R., 1947, *Supplement to Aust. Jl. Sc.*, **10**, No. 2.

LXXVIII. *Frictional Relaxation Oscillations.*

By B. R. DUDLEY and Professor H. W. SWIFT*.

[Received May 19, 1949.]

§1. INTRODUCTION.

WHEN a mass under elastic and possibly viscous restraint is subject to frictional traction from a steadily moving surface, it is known that under some conditions it may attain a position of stable equilibrium and that under others it may develop continuous steady oscillations of a relaxation type. It is characteristic of these oscillations that during part of each cycle the oscillating mass is dragged forward at the same speed as the moving surface, producing in fact the condition which has been realistically called "stick-slip".

The dynamics of this type of oscillation has been studied in the past, notably by Thomas (1930) and Blok (1940), for cases in which the coefficient of kinetic friction μ is independent of the speed of slip though it need not have the same value as the coefficient of static friction μ_0 . Thomas showed that in the absence of viscous damping stable oscillations could occur under these conditions, which would be simple harmonic if $\mu = \mu_0$, but of the "stick-slip" type if $\mu < \mu_0$. In the presence of viscous damping the simple harmonic type can no longer be sustained but stable "stick-slip" oscillations can be maintained provided that the damping is not excessive. Thomas showed that other conditions favourable to these continued oscillations are a small rubbing velocity and an appreciable margin ($\mu_0 - \mu$). Blok examined these vibrations in greater detail and defined the limiting conditions for "stick-slip" oscillations by means of a plotted relationship between dimensionless products for damping and rubbing speed. As it stands this relationship is valid only in a hypothetical case where the coefficient of kinetic friction $\mu = 0$, but Blok's analysis will be found to cover the more general case if the difference $\mu_0 - \mu$ is substituted for μ_0 in his results.

Practical interest in frictionally excited oscillations has been stimulated by recent attempts to measure the coefficient of friction between surfaces at various speeds of rubbing and to introduce certain physical explanations of "stick-slip" phenomena. Arising from this work certain observations and statements have appeared which cannot be critically examined in the light of any theory which assumes a constant kinetic frictional coefficient. For example, Bristow (1949) states that "for relaxation oscillations to be excited in the elastic system it is sufficient that the friction at the point of

* Communicated by the Authors.

contact with the moving surface decreases as velocity increases". He also states—"as would be expected from the theoretical considerations—that during the 'stick' part of the cycle there was, between the two surfaces, a relative movement which increased in velocity, causing a reduction in the magnitude of the relaxation oscillations ('stick-slips') as the velocity of the moving surface increased. As the speed of the moving surface further increases, the speed of 'slip' during 'stick' eventually equals the speed of the moving surface and relaxation oscillations cease. On the speed being still further increased, quasi-sinusoidal oscillations are maintained in the system by virtue of the decrease of friction with increase of velocity, as with the relaxation oscillations. This means that the motion of moving parts in a mechanism may consist of either relaxation oscillations, or quasi-sinusoidal oscillations, but not both, impressed on the general motion of the part".

In the present paper an attempt is made to examine the dynamics of relaxation oscillations with a frictional coefficient varying with rubbing speed, and to ascertain to what extent observations regarding such oscillations can be explained in terms of mechanics as applied to the accepted conditions of operation.

§2. PROCEDURE.

The dynamical system which we shall examine is shown in fig. 1, and the mass acceleration can be written

$$M\alpha = F - Sx - cv, \quad \dots \dots \dots (1)$$

where Sx is the elastic restraint, cv the viscous traction and $F = \mu P$ the frictional force transmitted by the rubbing surface moving with velocity u and under normal pressure P . We shall regard F as an empirical function of the rubbing speed $(u-v)$ having a limiting value $F_0 = \mu_0 P$ when $v = u$ and "stick" conditions are operative. Owing to the nature of the quantity F no functional solution of the equation is possible, but if this is written in the form $-v \frac{dv}{dx} \sqrt{(S/M)} = x \sqrt{(S/M)} - (F - cv) / \sqrt{(SM)} = \Delta$, say, it can be solved in any particular case, with a given F/v characteristic, by means of a graphical construction which develops the relationship between displacement x and velocity v .

In fig. 2 suppose that curve $F_0 F$ shows the relationship between the frictional criterion $F / \sqrt{(SM)}$ and the rubbing speed $(u-v)$, with U as origin, and the chained curve ACB the consequential relationship between $(F - cv) / \sqrt{(SM)}$ and the oscillating speed v , with O as origin. Then if the point P in the figure represents any momentary condition, so that $ON = v$ and $PN = x \sqrt{(S/M)}$ to the same scale as $F / \sqrt{(SM)}$, it follows that $PM = x \sqrt{(S/M)} - (F - cv) / \sqrt{(SM)} = \Delta$ and $PM/QM = \Delta/v = -dv/dx \sqrt{(S/M)}$. Hence, provided that the quantity $x \sqrt{(S/M)}$, which has the dimensions of velocity, is plotted to the same scale as v , the line PQ will be the normal at P to the curve relating $x \sqrt{(S/M)}$ to v . This curve may therefore be developed by drawing successive tangential elements of the type indicated at P in the figure, each being perpendicular to the appropriate line QP .

long as $v=u$, static friction will prevail and may have any value between the limits F_0 and $F'_0 = -F_0$, so that the operative point may lie anywhere within the range AA' .

If the locus for any motion meets the line AA' between these two limiting points, as indicated for example at K , then static friction will be sufficient to prevent any further change in velocity until the displacement of the oscillating mass, in conformity with that of the tractive surface, reaches the value corresponding to A . Beyond this point the elastic restraint predominates, the forward motion is retarded and kinetic friction becomes operative again. During progress along the line KA the body "sticks" to the tractive surface and any period of "stick" always extends to and ends at the point A . Hence in order to ascertain whether any given conditions will give rise to a "stick-slip" regime it is convenient to take A as the initial point from which to develop the locus. If this locus meets the line AA' between these two points the cycle will close at A and repeat itself indefinitely. If on the other hand the locus fails to intersect AA' "sticking" will not occur and the oscillation will start to damp out. If the locus intersects AA' produced, then the characteristic $A'B'$ comes into operation during the period of "over-run" until the locus meets or re-crosses the critical line AA' .

It is worthy of note that for given numerical conditions in equation (1) the locus defined by any point P is unique, and it follows that under these conditions no two loci can cross.

Assuming that a stable oscillation is sustained, the amplitude of this oscillation will be defined by the limiting values of $x\sqrt{(S/M)}$ on the locus, and these will of course occur where the locus crosses the prime axis $v=0$.

The labour of exploring the whole field of variations in the several quantities which jointly control a type of motion can often be reduced by application of the principle of dynamical similarity. In the present case this principle is of little avail. In the first place the coefficient of friction which controls the tractive force F is by hypothesis an empirical quantity determined solely by the rubbing speed $(u-v)$ as an independent variable free of any dimensionless product. In these conditions dynamical similarity can only exist between motions under a single tractive speed u .

Now the basic equation can be written in the dimensionless form :

$$(u/x_0)\sqrt{(M/S)} \frac{(v/u) d(v/u)}{d(x/x_0)} + (x/x_0)x_0/u\sqrt{(S/M)} - (F/F_0)F_0/u\sqrt{(SM)} + (v/u)c/u\sqrt{(SM)} = 0,$$

where x_0 is a displacement parameter, and F/F_0 is an empirical function of $(u-v)$ and therefore of v/u for a fixed u .

For dynamical similarity it is necessary not only that u shall be constant but also $F_0/\sqrt{(SM)}$ and $c/\sqrt{(SM)}$, which in effect set the field of fig. 2. The further requirement that $x_0\sqrt{(S/M)}$ shall be constant merely identifies a particular oscillation in that field. The disappointing fact is that the effect of changes in $F_0/\sqrt{(SM)}$, which is an important and wide-ranged variable, cannot be predicted by generalization but must be examined independently.

§3. RESULTS.

(a) *Modes of Vibration.*

In order to study frictional relaxation vibrations in a realistic way it is necessary to assume some appropriate form for the characteristic relationship between the frictional coefficient and rubbing speed. Such experimental evidence as is available suggests that a curve similar to F_0F in fig. 2 is representative of one type of characteristic (A), while a curve rising asymptotically from μ_0 to a larger value μ_1 at high rubbing speeds represents another (B). A rather less simple type (C) indicated in fig. 9, for which there appears to be experimental evidence under certain conditions, may also repay examination.

It is of course the AB curve, derived from the F_0F curve with allowance for damping, which finally determines the field of oscillations.

If this AB curve rises continuously with the rubbing speed ($u-v$), then it is easily seen that all vibrations eventually decay onto the point C, where steady operation ensues. A frictional characteristic of type (B) necessarily leads to an AB curve of this type. On the other hand, if the AB curve continually falls with increasing rubbing speed over the effective range, then small vibrations will tend to develop and larger ones to converge to the stable stick-slip oscillation characteristic of the field. Neither of these cases needs further detailed examination, though it should be noted that whereas the elastic displacement reached at C after the decay of vibrations is a true measure of the coefficient of solid friction at the relevant rubbing speed, the mean displacement $(x_1+x_2)/2$ of a stick-slip vibration is no measure of this coefficient.

For the present we shall consider frictional characteristics of type (A) which give rise to AB curves of sagging type having a minimum within the range of interest.

A general picture of the modes of vibration which can occur with a characteristic of this type may be obtained from fig. 3, in which the frictional criterion $F_0/\sqrt{(SM)}$ is given the same value as the speed u of the tractive surface and the damping ratio cu/F_0 is made 0.25.

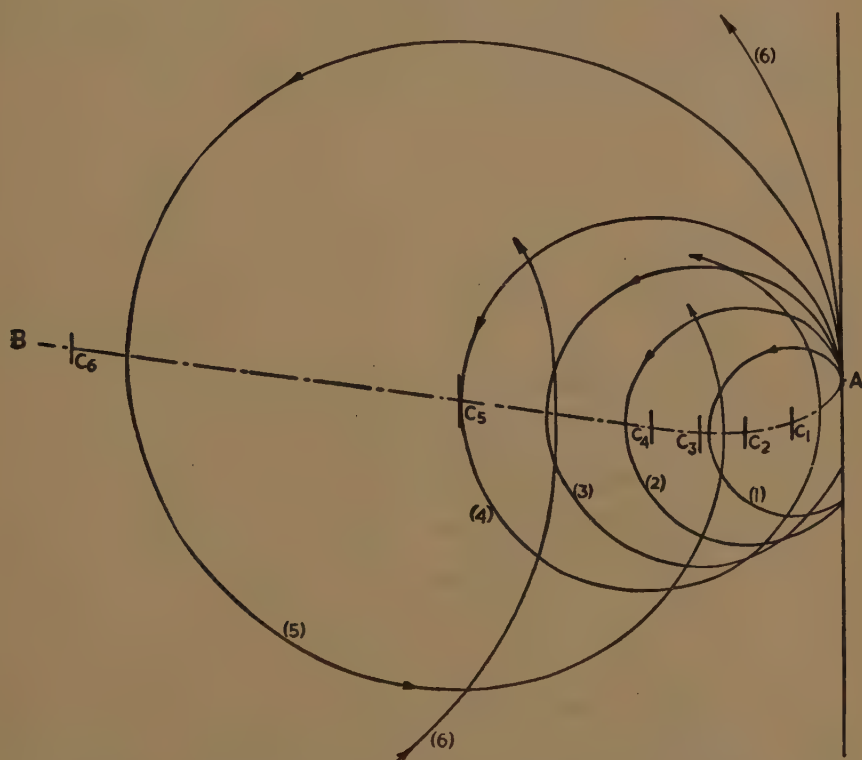
In this case the critical stick-slip curve AEK is stable (and of course unique) for all vibrations which meet the vertical through U. In the figure this curve has also been developed in the reverse direction from A, and it will be seen that it ultimately converges on a closed curve ZZ. It is clear that any vibration commencing at a point outside this curve ZZ will ultimately develop into the critical stick-slip cycle AEK, while it will be found that any vibration commencing within the curve ZZ is decadent onto the point C. Hence in the field represented by fig. 3, either stick-slip vibration or steady operation at C may occur according to initial conditions.

The general effect of viscous damping is illustrated by the curves in fig. 4, in which the critical loci have been plotted for various degrees of damping, measured by the respective intervals F_0A . It will be seen that increased damping reduces the range of "stick" and the amplitude of stick-slip, and enlarges the limiting curve ZZ within which decay will occur.

When the "stick" margin vanishes this curve ZZ coincides with the limiting stick-slip locus. With any greater damping, stick-slip conditions are unstable and all vibrations will decay onto the point C.

It is clear that small vibrations about C will tend to decay if the AB curve (fig. 2) is rising towards B at C, while they will tend to develop into stick-slip vibrations if this curve is falling. In this latter case the curve ZZ degenerates and steady operation at C cannot be attained.

Fig. 5.

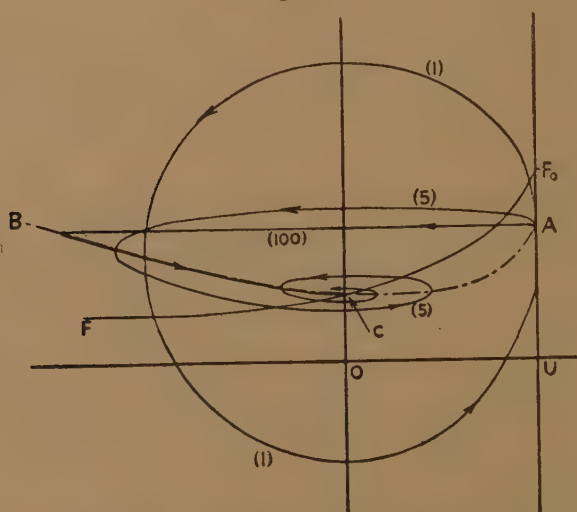


The effect of changes in tractive speed with a given damping ratio is illustrated by the curves in fig. 5. It will be seen that with the prescribed frictional characteristic slow speeds are conducive to stick-slip while at higher speeds the vibrations decay. It is notable, however, that as the speed is increased within the stick-slip range, the amplitude of the vibrations become steadily greater up to the brink of decadence, so that discontinuity occurs at this point and it has the features of a true "critical speed".

The effect of changes in the interfacial pressure can most conveniently be examined by assuming corresponding and proportional changes in the damping and elastic restoring forces; physically though perhaps less realistically this is tantamount to varying the mass M in equation (1).

In fig. 6 the three curves correspond to cases in which $F_0/u\sqrt{(SM)}=1, 5$ and 100 respectively, while $cu/F_0=0.3$ in each case. It will be clear from the figure that with the prescribed frictional characteristic and speed, stick-slip can occur at lower pressures, but at higher pressures the vibrations will decay, with or without an initial stick-slip according to initial conditions, and the locus may converge on C aperiodically. At high values of the ratio $F_0/u\sqrt{(SM)}$ the form of the vibration locus while it persists is closely related to that of the AB curve, and inspection of conditions in the region of C will show that the mode of vibration is very sensitive to the form and position of the trough of the AB curve in relation to the point C. The position of this trough to the right or left of point C determines whether the AB curve will rise or fall at this point and therefore whether small oscillations about C will decay or develop.

Fig. 6.



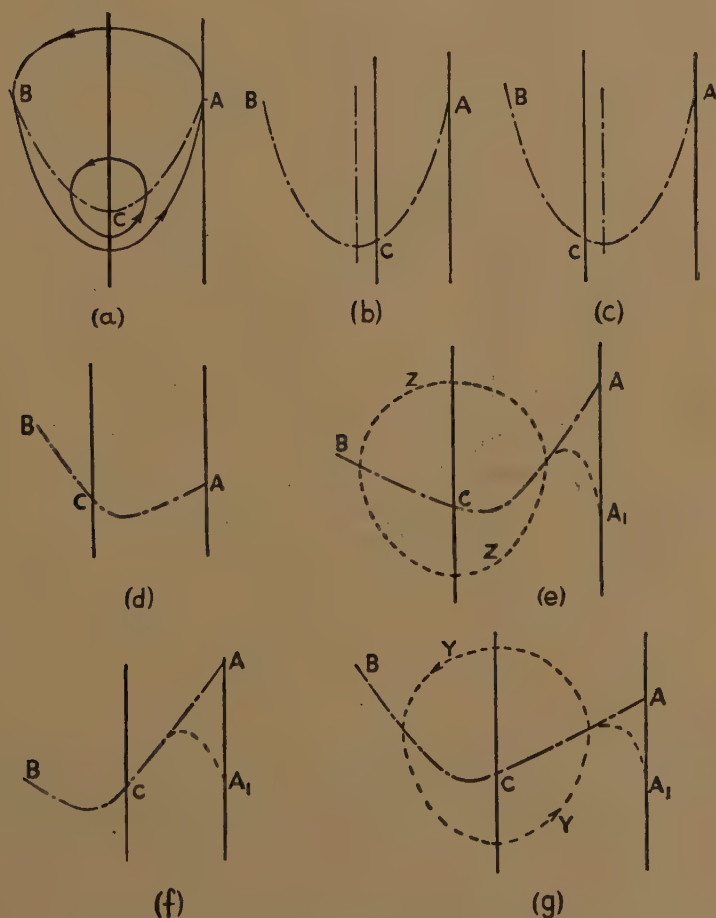
A number of conformations of the AB curve trough have been examined and the results can be summarized by reference to fig. 7. In this figure, diagrams (a)–(c) refer to AB curves which are assumed symmetrical about a vertical axis. In case (a), where the curve is symmetrical about the prime axis $v=0$, symmetrical oscillations may occur of the type indicated in the figure. True stick-slip is unstable but the limiting cycle through A can theoretically persist, as can any of the smaller cycles, though none is truly stable. In case (b) vibrations always develop to stick-slip, while in case (c) they always decay to C. If the AB curve is unsymmetrical, with its trough to the right of C as in cases (d), (e), then small oscillations will of course decay. In case (d) larger oscillations will also decay, but in case (e) they may develop into stick-slip if they commence outside a limiting curve ZZ.

Unsymmetrical curves such as (f), (g) with troughs to the left of C, lead to growth of small oscillations. This growth persists up to stick-slip.

conditions in case (f) but in case (g) stick-slip is unstable and decay commences. Hence in this case there is some limiting curve YY towards which all oscillations will either converge or develop according to whether they are initially larger or smaller, and this curve represents stable oscillations, which can be described as "quasi-sinusoidal" in the sense that they have no element of "stick".

It is possible that these oscillations are in fact those described by Bristow, since a reduction of the tractive speed in diagram (g) would give rise to

Fig. 7.



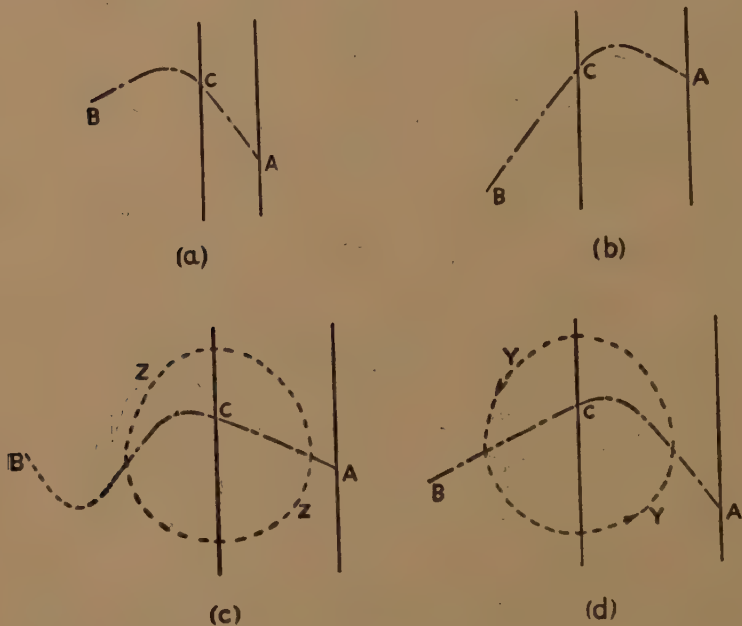
stick-slip conditions while an increase would lead to the conditions of diagram (d) and therefore to decay. It should be noted that these oscillations have in general no symmetry about C and the mean displacement is not therefore a true measure of the intrinsic friction at C.

Friction characteristics of the humped type (C) may either modify the results obtained with the sagging type (A) if the tractive speed is large

compared with the speed for maximum friction, or require separate consideration if the tractive speed is comparatively small. In fig. 7 cases (e) and (f), a fall to A_1 , as suggested by broken lines, might render the stick-slip conditions unstable and lead to stable "quasi-sinusoidal" oscillations instead. And in case (g) a similar fall would assist the decay from stick-slip towards the "quasi-sinusoidal" oscillations.

Cases where the tractive speed is comparable with that for maximum friction are shown in fig. 8. In case (a) vibrations will decay, while in case (b) they will develop. In case (c) small vibrations will decay but larger ones may develop to stick-slip, or possibly to intermediate stable

Fig. 8.



oscillations if the characteristic rises again (due to viscous damping) in the effective region. In case (d) small vibrations will grow while larger ones will tend to decay, so that the conditions are generally similar to those of case (g) in fig. 7 and stable "quasi-sinusoidal" oscillations of a definite form will result.

An interesting case is shown in fig. 9, which illustrates in a single field both the YY curve for stable "quasi-sinusoidal" oscillations and the ZZ curve defining the unstable limit between growth and decay. The trends of loci in the various zones are indicated by arrows.

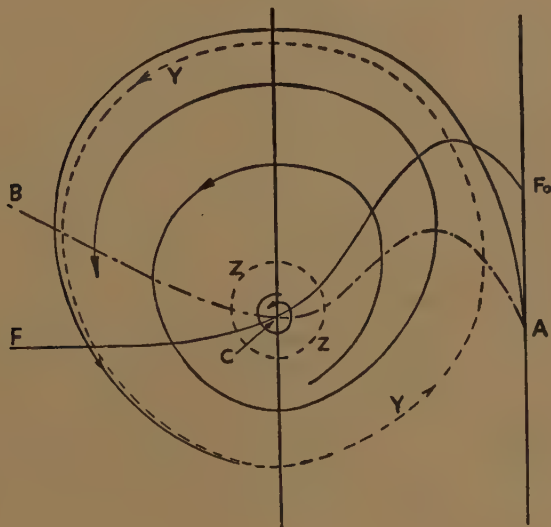
The comparative discussion of frictional vibrations has been confined to those initiated in such a way as to follow a locus which either damps out or intersects the vertical AA' (fig. 2) in the "stick" region between these two points. In practice initial conditions will as a rule correspond to one of the

two points O or U, and the vibrations will then fall within the scope of this discussion. If initial conditions are such that the locus intersects AA' produced, so that over-running takes place, then it will be found that with normal forms of AB curve the locus rapidly converges into the stick-slip range and its subsequent history falls within our discussion. It is possible however to prescribe a form of AB curve of the C type with little or no damping, which would permit stable over-running vibrations of a certain amplitude to develop and persist, if certain abnormal initial conditions were imposed. But it is not thought necessary to consider such cases in detail in this paper.

(b) *Periodic Times.*

The periodic time of a frictional oscillation of one of the types discussed above cannot in general be read directly from the x/v locus, but it can be

Fig. 9.



obtained from a derived curve or by summation from derived data. For since $\delta t = \delta x/v$, it follows that

$$\delta(tu\sqrt{S/M}) = (u/v)\delta(x\sqrt{S/M}) \quad \dots \quad (a)$$

$$= -(u/\Delta)\delta v. \quad \dots \quad (b)$$

Increments in the time criterion can therefore be computed in terms of increments in either v or $x\sqrt{S/M}$ by using factors obtainable direct from the x/v locus.

Alternatively the relationship between $tu\sqrt{S/M}$ and either v or $x\sqrt{S/M}$ can be developed by simple graphical construction. For in fig. 2 it is easily seen that

$$\cot \angle RQM = u/\Delta = -\delta(tu\sqrt{S/M})/\delta v.$$

Hence QR is normal to the curve relating v with $tu\sqrt{(S/M)}$ at the ordinate v , and the curve showing $tu\sqrt{(S/M)}$ in terms of v or alternatively of $x\sqrt{(S/M)}$ can be developed by obvious methods. This dual relationship proves convenient in operation because one or other tends to become indeterminate near turning points in the x/v locus.

Another potential basis for computing the periodic time is provided by the relation

$$\delta(t\sqrt{(S/M)}) = \delta(x\sqrt{(S/M)})/v = -\delta v/\Delta = \delta s/\sqrt{(v^2 + \Delta^2)} = \delta s/PQ,$$

where δs is an element of arc of the x/v locus. Now $\delta s/PQ = \delta\psi$, say, is the angle subtended at Q (on the prime axis) by the element of arc. The time corresponding to any part of the locus can therefore be measured by summing the elements $\delta\psi$ over this range. If Q were the centre of curvature of the locus at P this sum would be the resultant angle of flexure of the locus, which could be ascertained very easily for the arc corresponding to "slip". And since the time of "stick" is obtained simply by dividing the displacement along the AA' line by the "stick" speed u , the total periodic time could be determined with very little labour.

In general the point Q is not of course the centre of curvature, but so long as the damping is not too rapid, it is found that the evolute of the curvilinear part of the x/v locus does not depart far nor systematically from the prime axis, so that the error involved in the approximation is quite small and trends may be readily detected by its use. The approximation is embodied in the simple formula: $t\sqrt{(S/M)} = 2\pi - \theta + \tan \theta$, where θ is the angle ATK in fig. 2.

Representative values of the periodic time criterion $t\sqrt{(S/M)}$ are compared in the following table:

Type of oscillation	Fig. and curve	Time criterion	
		Computed	Estimated
Stick-slip	4, (1)	6.59	6.64
Stick-slip	Not produced	6.41	6.40
Quasi-sinusoidal	9, YY	6.38	6.28
Decadent	6, (5)	4.08	6.28

In all cases of course the natural period of oscillation of mass M under spring restraint S would be given by $t\sqrt{(S/M)} = 2\pi$.

If the periodic times are estimated for the loci in fig. 4, it will be found that these times fall as the damping increases from about 7.2 in curve (0) to 6.38 in curve (3), and 6.28 of course at incipient decadence. Similarly, in fig. 5 they fall as the speed increases, from about 9.5 in curve (1) to 6.6 in curve (3) and again 6.28 at incipient decadence. In general it is

clear that the periodic time of stick-slip oscillations will be greater than the natural period for the mass and elastic restraint, but will approach this value as the period of stick is reduced, and will reach it at the "critical speed". The period of "quasi-sinusoidal" oscillations is substantially the same as for free oscillations.

SUMMARY AND CONCLUSIONS.

A graphical method has been developed and applied to solve the equation of motion for frictional relaxation oscillations in cases where the intrinsic frictional coefficient varies with the speed of rubbing. The solution takes the form of a curve relating a displacement parameter with the speed of vibration.

A number of representative cases are examined, with frictional characteristics of types justified by experiment. It is shown that under certain frictional and velocity conditions stick-slip oscillations of a definite form will develop and persist, while under others oscillations will decay, and in some cases either stick-slip or decay can occur according to the initial displacement. As a rule slow tractive speeds lead to stick-slip oscillations and these increase in amplitude with the speed. At a certain critical speed however, stick-slip becomes unstable and the oscillations die out, leaving a steady displacement corresponding to the friction intrinsic to the tractive speed.

Under certain combinations of frictional characteristic and damping, stable "quasi-sinusoidal" oscillations may occur at tractive speeds intermediate between stick-slip and decay. These oscillations are not symmetrical about the position of static equilibrium.

The period of "critical" stick-slip is the same as of "quasi-sinusoidal" oscillations approximately the same as for free elastic oscillations.

In the range of stick-slip, the lower the tractive speed and the damping factor the greater is the periodic time.

REFERENCES.

- BLOK, H., 1940, *S.A.E. Journal*, **46**, 54.
BRISTOW, J. R., 1949, Paper presented to the Institution of Mechanical Engineers, 1st April 1949.
THOMAS, S., 1930, *Phil. Mag.* [7], **9**, 329.

LXXIX. *Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.*—Part I. *Observations with Photographic Plates exposed at an altitude of 11,000 feet.*

By Miss R. H. BROWN, U. CAMERINI, P. H. FOWLER, H. HEITLER,
D. T. KING and C. F. POWELL,

The H. H. Wills Physical Laboratory, University of Bristol*.

[Received June 15, 1949.]

[Plates VI.—XIII.]

SUMMARY

In this paper a phenomenological description is given of the results of observations made with "electron-sensitive" emulsions exposed at a height of 11,000 feet. It is shown that approximately half the nuclear disintegrations in which many charged particles are emitted are produced by particles of charge $|e|$, moving at relativistic velocities. The other half of these "stars" must be attributed to a neutral radiation. The directions of motion of most of the "primary" particles are inclined to the vertical at angles less than 40° . The particles appear to be able to pass through nuclei, imparting energy to the nucleons so that the nuclei evaporate, and having a 50 per cent probability of suffering a change of charge $|e|$ in each encounter. If the energy of the incident particle is sufficiently great, "showers" of charged particles are produced. Photo-micrographs are given of characteristic examples of "stars" of different types.

INTRODUCTION.

DURING the past few years, experiments with expansion chambers and counters have shown that the passage through matter of cosmic-ray particles of great energy leads to the production of "showers" of penetrating particles. We have in mind the numerous communications by Wataghin, Janossy, Rochester and Butler, Fretter, George and Jason, and other workers. It has been assumed by several authors that these events are due to the penetration of fast protons and neutrons through nuclei, and to the interaction of "primary" particles with the nucleons lying near their lines of motion. The difficulty of establishing the rest-mass of particles when they are moving at relativistic velocities is well known, and the nature of the shower particles has remained obscure. Further, it has not been clear whether the particles of a "shower" are commonly produced singly, in a succession of interactions of the primary particle with the nucleons along its path—the so-called "plural" production—or whether several particles can be created as a result of a single nucleon-nucleon collision—"multiple" production.

We have recently observed more than two hundred nuclear explosions, produced in photographic emulsions exposed to the cosmic radiation, which are accompanied by the emission of "showers" of fast particles

* Communicated by Professor C. F. Powell, F.R.S.

of low specific ionization (Brown *et al.*, 1949). These events appear to be identical in type with those which give rise to the "penetrating showers" observed in experiments with counters and expansion chambers. Because of the particular advantages of the photographic method we have been able to study certain features of these events in detail. The observations are complementary to those made by other methods. They give information about the nuclear transmutations produced by particles with an energy more than a hundred times greater than those which can be generated by the great machines now in operation. In this paper we describe the results obtained in an examination of plates exposed on the Jungfraujoch high altitude station, at 11,000 feet. In Part II., which is to follow, an account will be given of similar observations made at great altitudes by high-flying free balloons; and in Part III., the experimental results will be analysed and discussed.

EXPERIMENTAL.

The present observations were made with Kodak NT4 plates, with emulsions 200μ thick, or 400μ thick. The plates were processed in Ilford I.D. 19 developer, using a temperature cycle to secure uniformity of development (Dilworth *et al.* 1948). An examination of long rectilinear tracks, produced by particles of great energy, showed that even the thicker emulsions are free from distortion, except in regions less than 3 mm. from the edges of the plates.

TABLE I.
Details of Exposure.

Series	Place of exposure	Duration of exposure	Grain density g_{\min}	Matter above plates
E (1-35)	Jungfraujoch	14 days	420/mm.	7.5 cm. Pb, 1 cm. brass
F (1-24)	"	18 "	300/mm.	nothing
G (1-15)	"	42 "	—	"

The details of the exposures given to the plates are shown in Table I. The different series were developed with different processing conditions, time of development, etc., and the values of the grain-density in the tracks of particles of charge $|e|$ with minimum ionization are included, for each series, in the Table I.

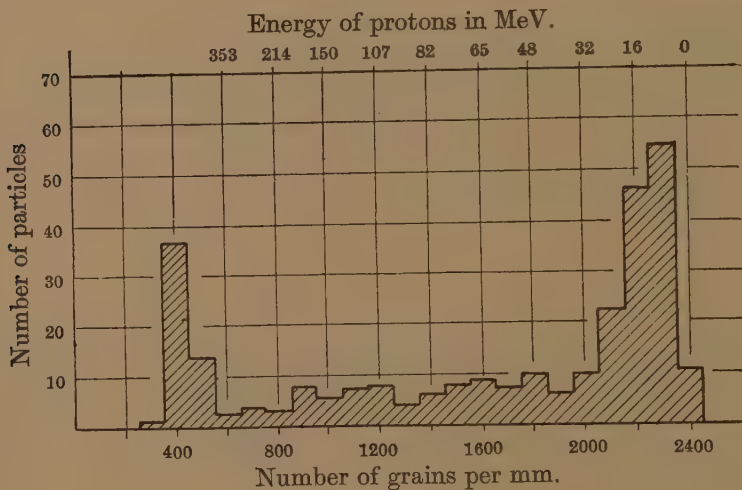
"FAST" AND "SLOW" CHARGED PARTICLES.

Among the "stars" produced in the emulsions, we have observed twelve hundred examples, with each of which there is associated at least one fast particle, with a specific ionization near the minimum value for charge $|e|$. The distribution in the values of the grain-density in the tracks of 300 particles, originating in "stars" from which at least seven heavily ionizing particles diverge—mostly protons and α -particles—is shown in fig. 1. The observations were confined to tracks with a length greater

than $300\ \mu$, and the results clearly display the peak in the distribution at a grain-density of 400 grains per mm., a value equal to that for isolated tracks of fast particles recorded in the same emulsion. It will be convenient to represent by the symbol i_{\min} , the minimum value of the ionization of a particle of charge $|e|$, and by g_{\min} , the corresponding value of the grain-density in the track.

In fig. 1, there is a second maximum at a high grain density, of the order of 2400 grains per mm., due mainly to protons of relatively low energy, 15–40 MeV.—the tracks of which are made up of an almost continuous succession of grains—together with a continuous distribution between the two peaks. (The tracks with a grain density less than about 1200 grains per mm. would not have been recorded by the less sensitive emulsions employed until recently.) In what follows we shall apply the term “fast” particle to any one which produces a track with a grain-density

Fig. 1.



Histogram showing a typical distribution in the values of the grain-density in the tracks associated with stars. The measurements have been confined to tracks, of length greater than 300μ , which are associated with “stars” of seven or more branches.

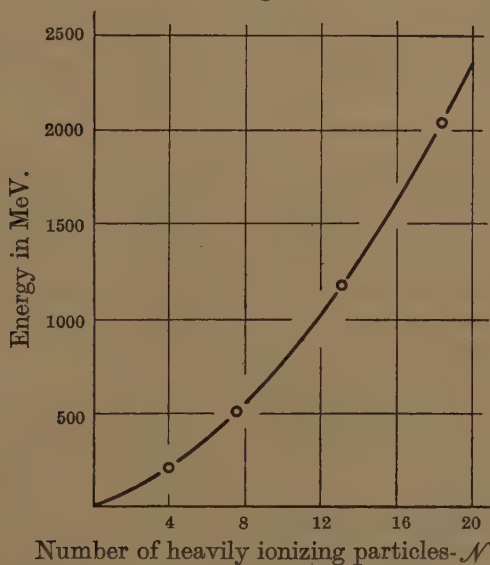
less than $1.5\ g_{\min}$, and shall refer to the corresponding tracks as “thin” tracks. The value of the ratio of the kinetic energy, E , of such a particle, to mc^2 , where m is the rest-mass, is greater than 2. We represent the number of more heavily ionizing particles in a star, the tracks of which have a grain-density greater than $1.5\ g_{\min}$, by the symbol \mathcal{N}_h . It will be convenient to distinguish between “grey” tracks, with a value of the grain-density between 1.5 and $5\ g_{\min}$, and “black” tracks with $g > 5\ g_{\min}$. The “grey” tracks are almost all due to particles of charge $|e|$; and the energy of the corresponding particles, if they are protons, lies in the interval from 25 to 330 MeV.

EFFICIENCY OF DETECTION OF FAST PARTICLES.

In interpreting the results it is important to know the efficiency, f , with which the tracks of "fast" particles can be distinguished in our particular experimental conditions. An estimate of this quantity can be made in the following way:—

The great majority of the π -particles, which stop in the emulsion and decay to produce μ -particles of energy 4.25 MeV., are positively charged. In about 10 per cent of these events, the μ -particle is also brought to the end of its range in the emulsion; and we may assume, tentatively, that it then invariably decays with the emission of a fast electron, the track of which has a grain-density equal to the minimum value, g_{\min} . A comparison of the observed proportion of the events of this type in which the decay electron is actually observed then gives a measure of f .

Fig. 2.



Average energy release in "stars" in which different numbers of heavily ionizing particles are emitted. The values of the energy correspond to that of the emitted nucleons and do not include that of any "shower" particles.

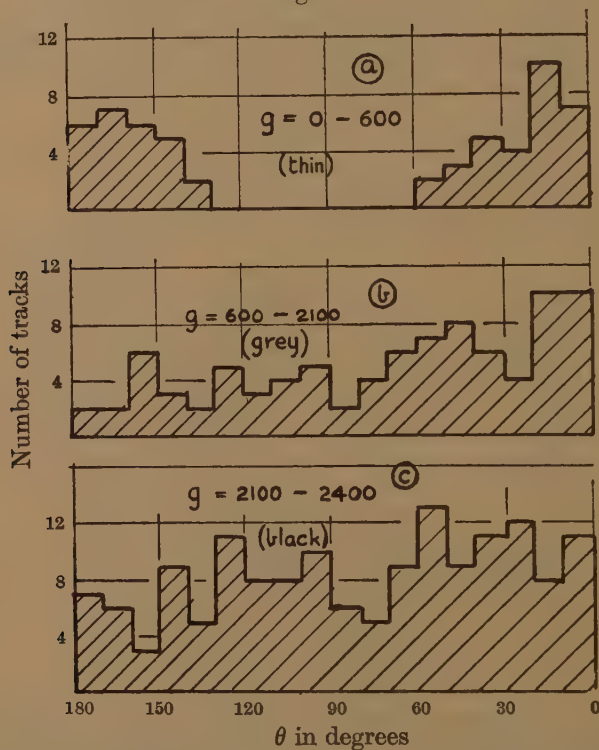
We have found 52 events showing the transformation $\pi \rightarrow \mu$, in the plates under examination, in which the μ -particle stops in the emulsion; and in 42 of these events, the track of the electron can be distinguished. We conclude that $f > 0.80$. This figure refers to the efficiency of detection of the tracks of particles of which both the points of origin and the directions of motion are distributed at random. We shall see, however, that the majority of fast particles associated with stars move in directions making angles less than 50° with the plane of the emulsion. The efficiency of detection of the track of a fast particle is greater, the smaller the angle at

which it "dips" relative to the plane of the emulsion. The observed value of f , deduced by observations on the decay electrons, therefore represents a lower limit to the true value for the fast particles associated with "stars."

ENERGY RELEASE CORRESPONDING TO "STARS" OF DIFFERENT SIZES.

A second quantity which is of importance in interpreting the results is the average release of energy represented by "stars" in which there are different numbers of "grey" and "black" tracks; *i. e.*, those with a

Fig. 3.



Distribution of the values of θ for "thin", "grey" and "black" tracks.

In this case, the observations have been restricted to tracks, of length greater than 300μ , associated with stars of seven or more branches ($\mathcal{N}_h \geq 7$).

specific ionization greater than $1.5 i_{\min}$. We have determined values of this quantity by the methods described in the appendix, and the results are represented in fig. 2. The values appear to be in reasonable accord with the results of experiments at Berkeley on the disintegrations produced by artificially accelerated deuterons and α -particles of known energy; and with observations on the disintegrations produced by π^- -particles brought to rest in the emulsion, if it is assumed that all the energy corresponding to the rest-mass of such a particle is shared among

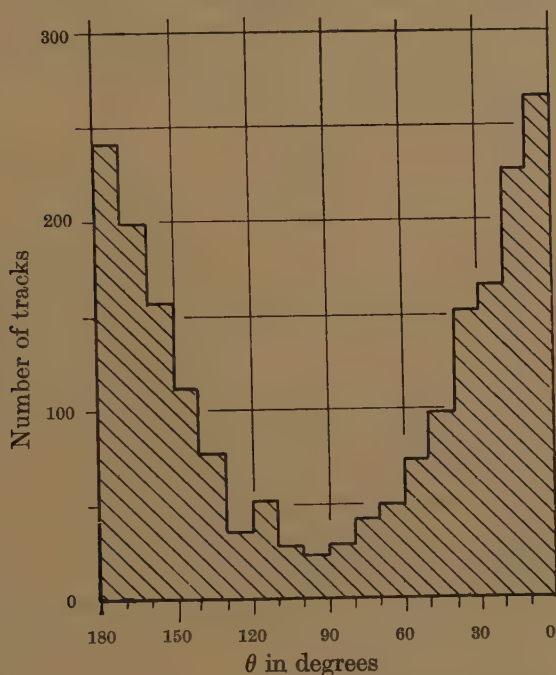
the nucleons of the nucleus which it disintegrates. The curve drawn through the points in fig. 2 is calculated from the empirical relation :

$$E(\text{MeV.})=37\mathcal{N}+4\mathcal{N}^2.$$

ORIENTATION OF THE TRACKS OF FAST PARTICLES ASSOCIATED WITH "STARS".

During the exposure the plates were arranged in a defined orientation with the emulsion lying in a vertical plane, and it is therefore possible to determine the direction of motion of a particle, in its passage through the

Fig. 4.



Distribution in the values of θ , of the tracks of all fast particles associated with "stars". θ is the inclination to the vertical of the projection of a track on the plane of the emulsion. 0° corresponds to the direction vertically downwards from the "star".

emulsion, relative to the vertical. For simplicity, we have usually chosen to measure only the direction, θ , made by the projection of a track on the plane of the emulsion, relative to the vertical. A histogram showing the distribution of the observed values of θ for all "thin" tracks of length greater than 300μ , taken from "stars" with which at least seven heavily ionizing particles are associated, $\mathcal{N}_h \geq 7$, is shown in fig. 3(a). In fig. 3, and other diagrams of the same type, 0° corresponds to the direction vertically downwards from the centre of disintegration.

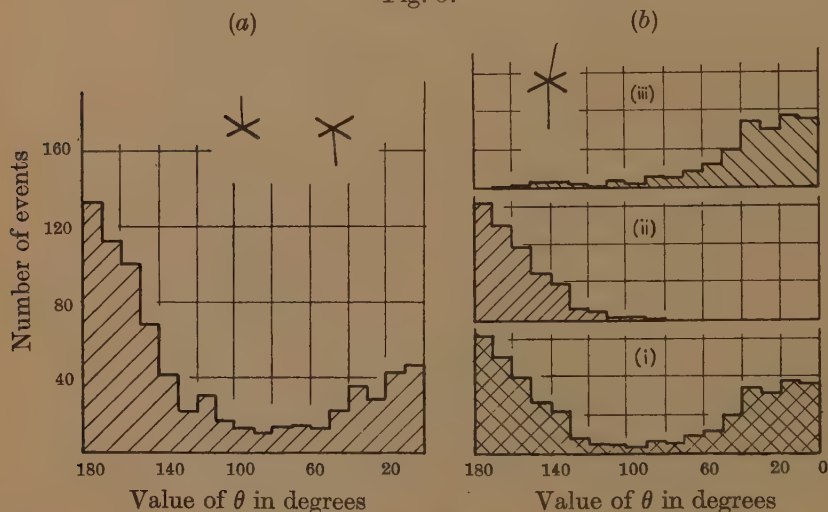
It will be seen from fig. 3(a) that the lines of motion of the fast particles in events of this class, are inclined at angles less than 40° to the vertical.

The results suggest that the great majority of tracks "above" the stars are due to particles which approached a nucleus and produced its disintegration. This conclusion receives decisive support from other observations described in a later paragraph.

Fig. 3(c), shows the corresponding distribution in the values of θ for the heavily ionizing particles which produce "black" tracks, $g > 2100$ grains per mm. It will be seen that for these particles, in contrast with the results for the fast particles, there is a nearly isotropic distribution of the directions of emission. The results of similar observations on the tracks with intermediate values of the grain-density are given in fig. 3(b), and display a well-marked tendency to be ejected in the downward direction; but they are not so sharply collimated as the fast particles.

Fig. 4 shows the distribution in θ of all thin tracks, irrespective of their length, associated with stars of $\mathcal{N}_n \geq 3$.

Fig. 5.



Distribution in the values of θ for the tracks of fast particles; (a) for events with which one fast particle is associated; (b) two fast particles. In fig. 5(b), the original observations, (i) have been analysed to show the distribution in the values of θ for incoming particles, (ii); and outgoing particles, (iii).

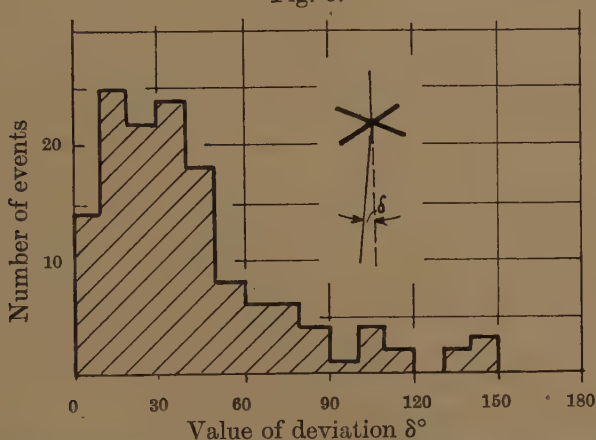
We have made a similar analysis of those stars with which one or two fast particles are associated. The lines of motion of the fast particles are again found to be confined, in the majority of cases, to directions making only small angles with the vertical; see fig. 5. Here again, the results strongly suggest that the track on the upper side of the star is produced by a "primary" particle.

In some cases, in which two fast particles are associated with a star, they both occur in the upper hemisphere. If we assume that, in all cases, the primary particle is the one producing the track making the smallest angle

with the vertical, we can determine the angular distribution of the secondary fast particles—see fig. 5 (b) (iii). Fig. 5 (b) (ii), shows the corresponding distribution of the primary particles which, because of the above convention, necessarily reaches zero for $\theta \geq 90^\circ$.

A histogram showing the distribution of the observed values of the difference in the directions of motion of the two fast particles associated with each event of this class is shown in fig. 6. In making these particular observations we have examined the spacial orientation of each track, taking account of the angle of “dip” in the emulsion. The true deviation, δ , has thus been determined and not merely its projection on the surface of the emulsion. It will be seen that the difference in direction is usually less than 50° , but that much larger values are sometimes found.

Fig. 6.



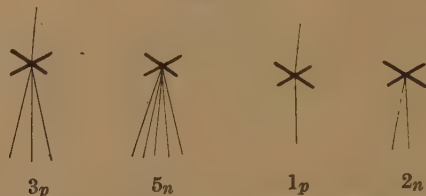
Histogram showing the distribution of the angle between the directions of motion of the two fast particles in the events of type 1_p .

NOMENCLATURE.

On the basis of the evidence presented in the previous paragraphs it appeared reasonable to classify the “stars” with which fast particles are associated, in the following way. A star is characterized according to the number, n_s , of fast particles produced in the nuclear interaction—*i. e.*, the number of “shower” particles. If the track of a fast particle is also associated with the star, which we can reasonably assume, because of its direction of motion, to be that of an incident particle which produced the disintegration, then the suffix “ p ” is added to the number n_s . If no such track is visible, we assume the event to have been produced by some form of neutral radiation, and we then add the suffix “ n ”. This interpretation will be incorrect if the particle producing the disintegration was of charge $2e$ or greater; or if the primary particle, of charge e , had a velocity below that corresponding to minimum ionization. Simple illustrations of this

nomenclature are shown in fig. 7, where different classes of stars are represented schematically. For the further description of an event it is sometimes convenient to specify the number of heavily ionizing particles, \mathcal{N}_h , ($g > 1.5 g_{\min}$) which accompany it.

Fig. 7.



Schematic representation of stars of different types.

Although we shall sometimes be in error in assuming a particular track to be that of a particle which approached the nucleus, the degree of collimation at 11,000 feet is sufficiently well-marked to make it certain that the resulting errors are trivial in their bearing on the results of the present paper.

STARS OF TYPES 1_p AND 1_n .

Fig. 8(a) shows the relative frequency of those events of types 1_n and 1_p , in which the production of a single fast particle is accompanied by the emission of different numbers, \mathcal{N}_h , of "slow" charged particles, $g > 1.5 g_{\min}$. By using the results given in fig. 2, a scale showing the average value of the total energy released among the nucleons, for the stars with different numbers of slow particles, has been included in the figure. Typical photo-micrographs of disintegrations of type 1_p are shown in Pl. VI.

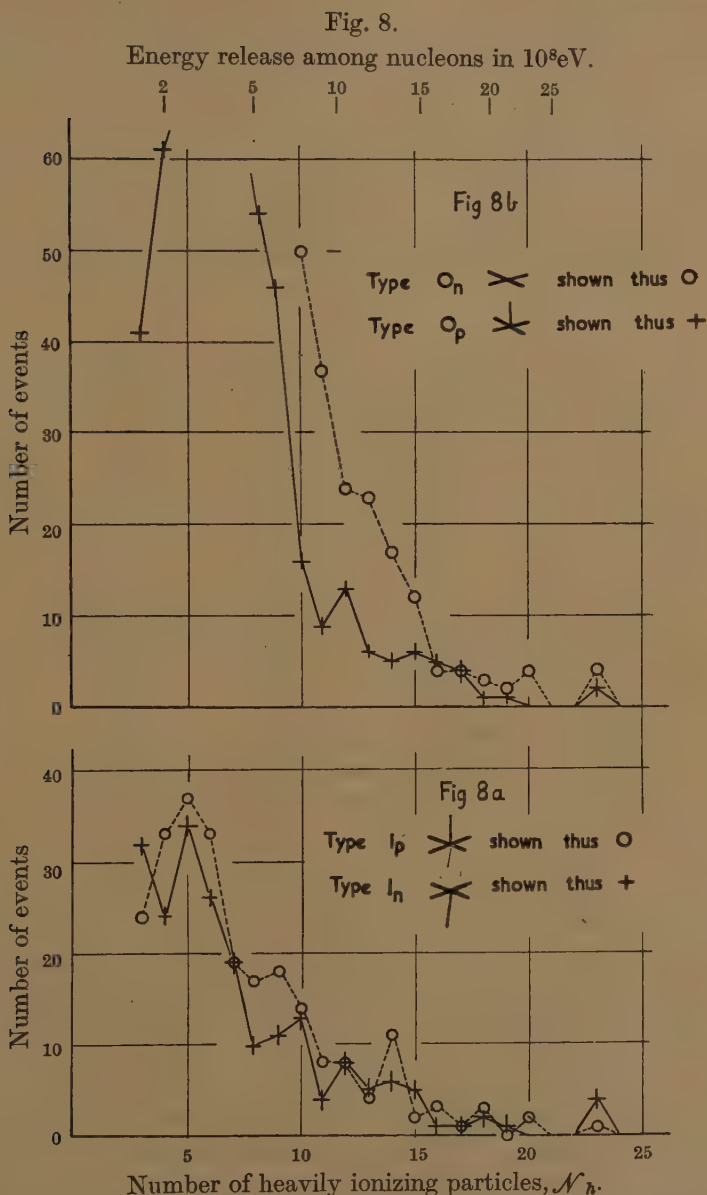
It will be seen from fig. 8(a) that the results for the two types of events display the same form of distribution, and that their frequency of occurrence is equal to within the statistical fluctuations. This striking result shows that the neutral and charged particles giving rise to disintegrations of these types are present with equal intensities in the cosmic radiation at 11,000 feet.

The distribution in the values of the energy release in stars of type 0_p , analogous to that for stars of types 1_n and 1_p given in fig. 8(a), is shown in fig. 8(b). A comparison of figs. 8(a) and 8(b) shows that for stars with many ejected charged particles, $\mathcal{N}_h > 12$, the frequencies of occurrence of the three types are equal to within the limits corresponding to statistical fluctuations; but that for smaller stars, the number of type 0_p is about twice that of either 1_n or 1_p . Stars of class 0_n are more frequent than those of any other classes for all values of N_h . For examples of stars of these classes, see Pls. VII. and VIII.

SHOWERS OF FAST PARTICLES.

The most striking phenomena observed in the present observations are the nuclear transmutations which are accompanied by the emission of

"showers" of fast particles, and mosaics of photo-micrographs of two examples are shown in Pls. IX. and X. The ratio of the grain-density



Observed frequencies with which stars, of types O_n , O_p , I_n and I_p , have different numbers of "branches", \mathcal{N}_h . For $\mathcal{N}_h > 10$, the types O_p , I_n and I_p occur with equal frequency. Types I_n and I_p are equally numerous for all values of \mathcal{N}_h .

of the fast particles produced in these two of events, to the value characteristic of minimum ionization for a particle of charge e , is shown

in Table II. The showers of fast particles, shown in Pls. IX. and X., are each accompanied by only four slow, heavy, charged particles ejected from the original nucleus in directions apparently orientated at random. In other events of a similar type, many heavy particles are associated with the shower, and two typical examples are represented in Pls. XII. and XIII. Pl. XIII. shows a facsimile drawing made with the aid of the projection microscope.

The distribution in the values of θ for the fast particles occurring in events with which three or more shower particles are associated, is shown in fig. 9. It will be seen that the form of the distribution is similar to that given in fig. 5(a). In this case also, therefore, the observations give strong

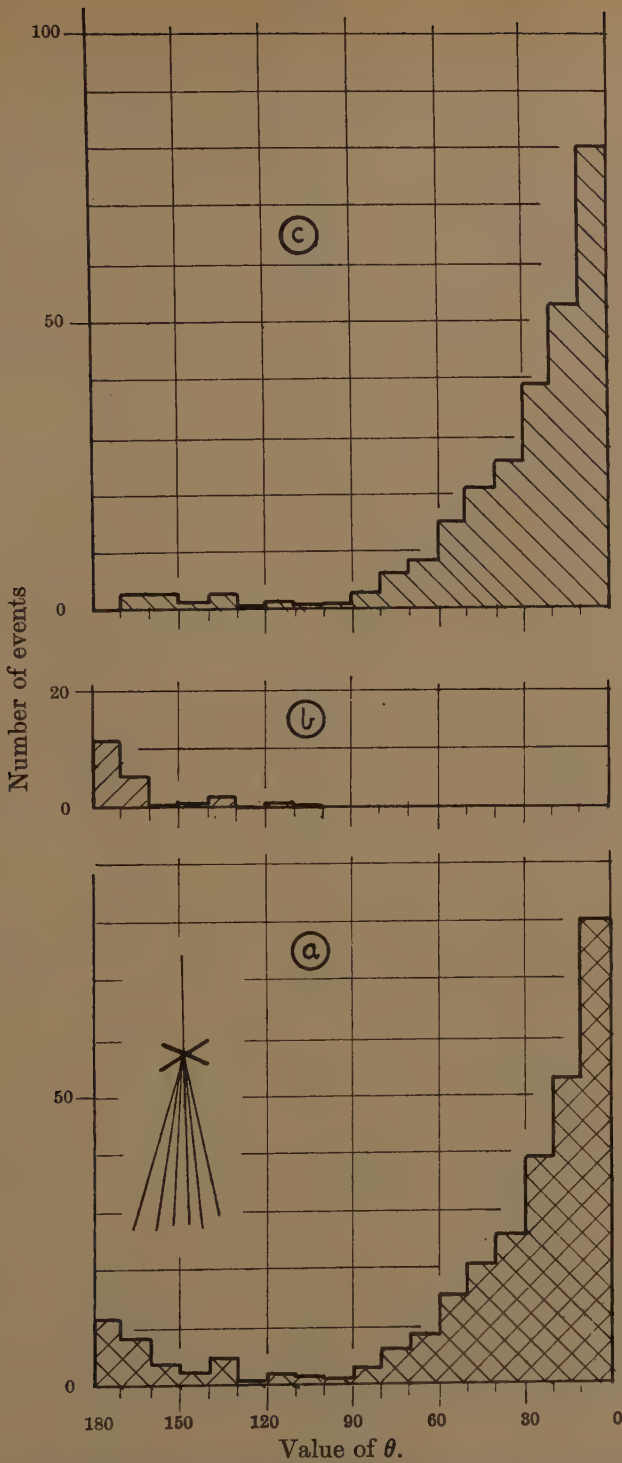
TABLE II.

Star no.	Track no.	Length of projection	Number of grains	cosine of angle of dip	g/g_{\min} .
KE 25	1	130	58.5	.836	.96
	2	631	267.5	.996	.95
	3	223	104	.934	1.09
	4	128	58	.832	.99
	5	1757	759.5	.999	1.07
	6	174	71	.897	.92
	P	297	110.5	.981	.87
KE 8	1	375	137.5	.988	.86
	2	262	89.5	.976	.79
	3	545	224	.996	.97
	4	98	52	.886	1.15
	5	252	137	.974	1.27
	6	560	242.5	.990	1.02
	7	81	31.5	.811	.79
	8	75	33.5	.788	.90
	9	74	75.5	.822	1.16
	10	220	87	.974	.92

support for the view that at least the majority of the tracks of fast particles "above" the "stars" are to be regarded as due to charged incident particles which produced the associated disintegrations. In order to confirm this view, we determined the spacial orientation of the tracks of all the fast particles in events of this type, and typical results are shown in fig. 10.

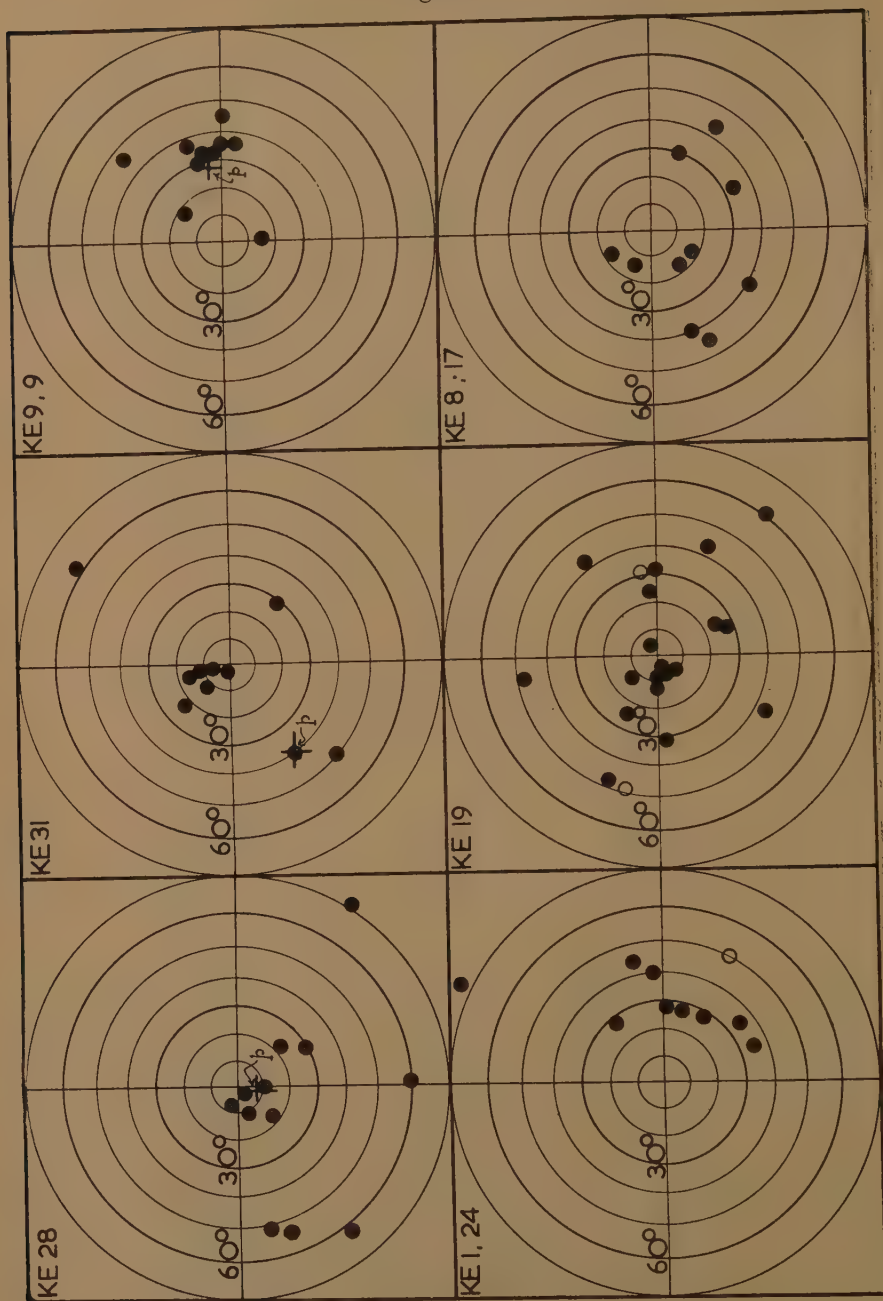
Each of the target patterns shown in fig. 10 represents the points of incidence of the "shower" particles, associated with a particular event, on the lower hemisphere. A full circle, ●, indicates the direction of motion of a "shower" particle; and a cross, +, that of an incident particle, assuming it to have approached the nucleus. In a few cases, a "shower" particle is ejected in a direction contained in the upper hemisphere. In this case its direction of motion is represented by an open circle, ○. The position of this point in the diagram corresponds to that of a particle moving in the same line of motion as the particles in question, but in the opposite direction.

Fig. 9.



Values of θ for tracks of fast particles associated with "showers" — $n_s \geq 5$.
 (a) shows the distribution for all tracks; (b) for tracks attributed to incoming particles; and (c) for "shower" particles.

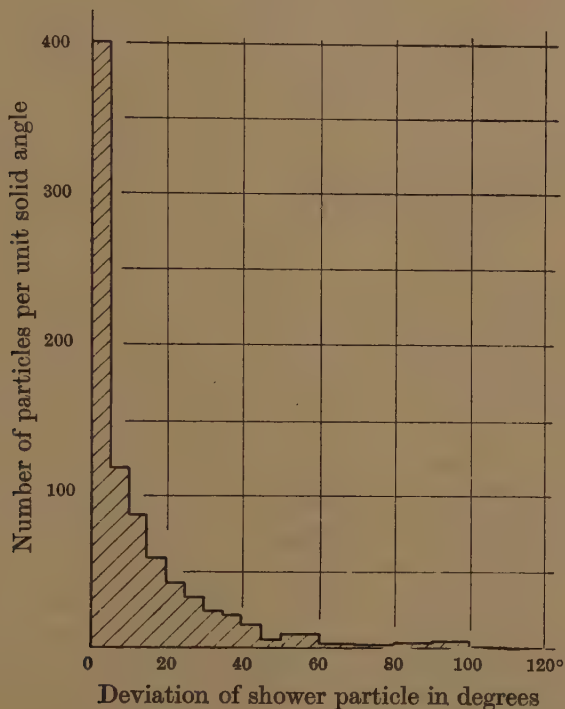
Fig. 10.



"Target" patterns showing typical cases of the orientation of the tracks of "shower" particles with respect to the vertical. For the events represented in the three left-hand patterns, the charged primary particle can be distinguished and the point corresponding to its direction of motion is indicated thus : +_p.

It will be seen from fig. 10 that in those events in which the track of an incident particle can be distinguished, its direction of motion lies close to the "centre of gravity" of the shower particles which it is assumed to have produced. In general, the association is particularly well marked in the case of showers of ten or twelve particles, especially when there is a concentrated "core" of particles in the shower. We regard this as decisive evidence that the single isolated particle is indeed the primary particle responsible for the disintegration, in at least the great majority of events of this type.

Fig. 11.



Distribution in the directions of motion of shower particles with respect to that of the primary particle.

Observations of the type represented in fig. 10 enable us to determine the distribution of the directions of motion of the shower particles with respect to the charged primary particles, and the result thus obtained is shown in fig. 11. This figure shows the number of particles per unit solid angle at different inclinations, δ , to the direction of motion of the charged primary.

In about half the cases in which showers of five or more charged particles are produced, no track with a value of the grain density corresponding to minimum ionization can be distinguished which can be attributed to an incident particle. We therefore conclude that in these events the particle producing the disintegration was electrically neutral.

TABLE III.—Number of heavily ionizing particles, N_h .

	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21-26	>27	Total
0n	1030	1063	741	357	186	103	65	50	37	24	23	17	12	4	4	3	2	4	4	..	3729
0p	41	61	92	91	71	54	46	16	9	13	6	5	6	5	4	1	1	0	2	..	525
1 ⁿ p	24	33	37	33	19	17	18	14	8	8	4	11	2	3	1	3	1	2	1	..	238
1 ^p p	32	24	34	26	19	10	11	13	4	8	5	6	5	1	1	2	1	..	4	2	208
2 ⁿ p	5	8	8	6	5	4	5	5	3	2	..	2	1	..	1	1	54
2 ^p p	2	7	6	6	7	6	5	3	4	5	5	2	..	1	1	3	1	69
3 ⁿ p	1	2	2	1	1	2	1	..	2	1	16
3 ^p p	1	2	4	..	4	4	1	1	..	1	1	1	21
4 ⁿ p	2	1	1	..	2	1	1	1	1	1	11
4 ^p p	1	1	1	1	..	1	1	2	1	2	13
5 ⁿ p	1	3
5 ^p p	1	2	..	1	2	1	7
6 ⁿ p	1	1	1	1	1	1	2
6 ^p p	6
7 ⁿ p	0
7 ^p p	..	1	1
8 ⁿ p	..	1	1	2
8 ^p p	1	1	..	2
9 ⁿ p	1	1	1
9 ^p p	1	1	4
10 ⁿ p	1	2	1	0
10 ^p p	1	1	2
11 ⁿ p	1	..	3
11 ^p p	2	0
12 ⁿ p	1
12 ^p p
13 ⁿ p	1	↓	↓
13 ^p p	17p	21n	..	2

Number of emitted "shower" particles, n_s .

To reduce the observed values to numbers per c.c. per day, multiply by $\sim 2.2 \times 10^{-3}$. The resulting numbers then give the estimated intensities of stars of different times in conditions in which there is no absorbing material above the plates.

PRODUCTION OF π -PARTICLES.

We have found examples, among stars of all classes, of disintegrations which are accompanied by the emission of slow mesons, and characteristic photographs of such events are shown in Pl. XI. In 70 per cent of the cases in which such a meson reaches the end of its range in the emulsion, it produces a disintegration with the emission of heavy charged particles. In no case have we observed such an ejected meson to decay with the emission of an electron. In previous papers from this laboratory, it has been shown that at least 95 per cent of the slow mesons emitted from stars, and which stop in the emulsion, are negatively charged. This result was attributed to the effect of the nuclear change, which will tend to cause positive particles to be emitted with greater kinetic energy than the negative, and thus cause them to escape from the emulsion except in rare cases. The present observations are consistent with the previous experiments in indicating that at least the great majority of the slow mesons emitted during nuclear explosions are π^- -particles; and the observations are consistent with the assumption that they are all of this type.

RELATIVE FREQUENCY OF PRODUCTION OF DIFFERENT NUMBERS, n_s , OF SHOWER PARTICLES.

Table III. shows the relative frequency of production of nuclear explosions at 11,000 feet, in which a given number of shower particles, n_s , is associated with the emission of a number, \mathcal{N}_h , of more heavily ionizing particles—grey and black tracks. Those types of events with which fast primary particles are associated are distinguished by the letter p ; otherwise by n .

The relative frequency with which different numbers of shower particles are produced, irrespective of the number of associated heavily ionizing particles, \mathcal{N}_h , is shown in fig. 12. The observations are consistent with the assumption of a continuous diminution of the frequency of occurrence of events with increasing values of n_s . There is some indication, however, of a rapid fall up to $n_s=6$, followed by a nearly constant frequency of occurrence of events with n_s between 7 and 12. Such an effect, if it exists, would appear to have a bearing on the problem of the multiplicity of production of "shower" particles in single nucleon-nucleon collisions, and it is therefore important to increase the statistical weight of the observations.

APPENDIX.

ENERGY RELEASE REPRESENTED BY "STARS" OF DIFFERENT NUMBERS OF PARTICLES.

In order to determine the average values of the release of energy among the nucleons, in nuclear "explosions" represented by "stars" with different numbers of branches, we first divided the observed events into four groups according to the value of \mathcal{N}_h ; $\mathcal{N}_h=4$ or 5; $\mathcal{N}_h=7$ or 8; $\mathcal{N}_h=12$ to 16.

inclusive; and $\mathcal{N}_n=16$ to 25 inclusive. For the events of each of these classes we distinguish three types of tracks:

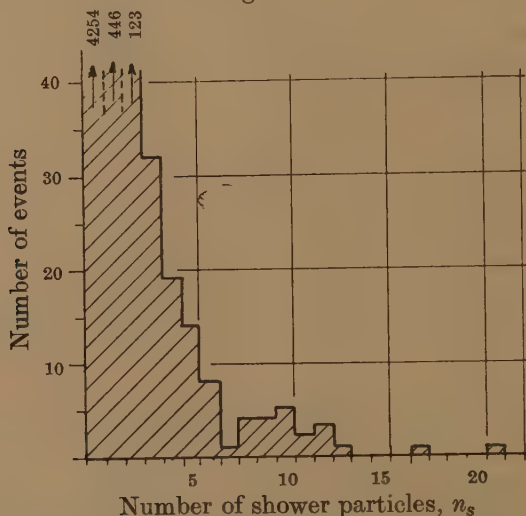
(a) short tracks, of length $< 70 \mu$, produced by particles which reached the end of their range in the emulsion;

(b) "black" tracks, $g > 2100$ grains per mm. produced by particles which left the emulsion, or which reached the end of their range after travelling a distance greater than 70μ ; and

(c) "grey" tracks. $2100 > g > 600$ grains per mm.

For each class of star, we can deduce the average number of particles per star which are emitted with a range less than 70μ . Many of these particles escape from the emulsion without coming to rest, but we can

Fig. 12.



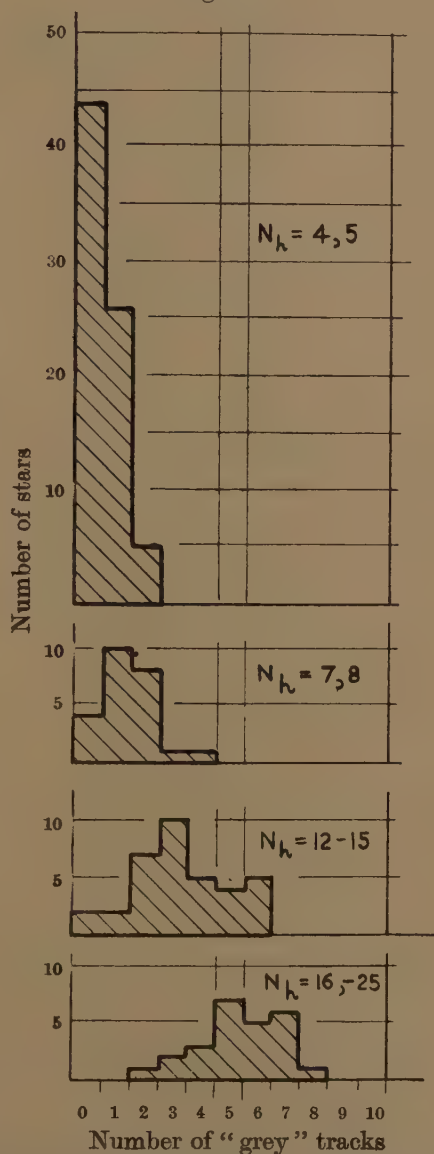
Observed frequency of nuclear explosions with which different numbers, n_s , of shower particles are associated.

determine the effect of this "escape" by a simple geometrical calculation; and the observed number which reach the end of their range in the emulsion, tracks of class (a), then allows us to calculate the total number we should expect to observe in an emulsion of indefinitely great thickness. We assume that most of these tracks are due to α -particles, but some of them will be produced by energetic nuclear fragments of greater mass, and others by protons of short range. We estimate that the average value of the energy required for the production of each of these particles is 15 MeV.; 10 MeV. for the kinetic energy of the particles, and 5 MeV. binding energy.

By similar observations on the tracks of type (b) we calculated the true numbers of "black" tracks produced by particles of range $> 70 \mu$. We assume that the great majority of these tracks are due to protons and deuterons, and that the production of each track requires an energy of 21 MeV., irrespective of the class of the "star" with which it is associated.

This assumption appears to be justified by the fact that the distribution in the values of the grain-density in the "black" tracks is observed to be independent of \mathcal{N}_h , and that the grain-density curves for stars of all

Fig. 13.



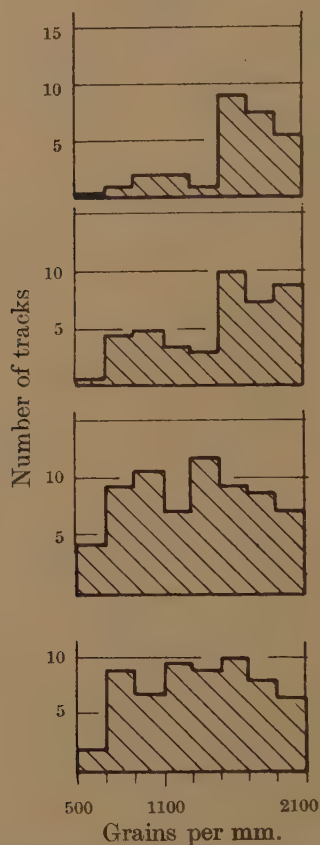
Histograms showing the distribution of the number of "grey" tracks for "stars" of different classes.

classes show the sharp drop, of the type represented in fig. 1, for values of g less than 2100 grains per mm. The emitted protons must be accompanied by neutrons, and we assume the latter to correspond to a release of energy

25 per cent greater than that of the protons, corresponding to the ratio of the numbers of neutrons and protons in the nuclei of silver bromide.

Most of the energy released in the nuclear explosions is carried by the particles which produce the "grey" tracks. The relative frequency with which different numbers of "grey" tracks are associated with "stars" of the four different classes, is shown in fig. 13. The spread in the values for any

Fig. 14.



Histograms showing the distribution in the values of the grain-density of the "grey" tracks associated with "stars" of different classes.

one class is only a little greater than that corresponding to a Poisson distribution. Because of the large fraction of the energy represented by the grey tracks, this result suggests that the total energy release in a particular star with a given value of \mathcal{N}_h is never widely different from the mean value for stars of its class, the fluctuations depending largely on whether the ejected fast nucleons of energy between 40 and 400 MeV., appear as neutrons or protons.

The distributions of the values of the grain-density in the "grey" tracks, for the four different classes of stars we have chosen to consider, are shown



Two examples of disintegrations of type 1_p , in which the impact of a "relativistic" particle with a nucleus is accompanied by the emission of a single "fast" particle. The small difference in the directions of motion of the "fast" particles in each of the events is typical of most of the disintegrations of this class. Both events are exceptional, however, in showing the tracks, " e ", of electrons with an energy of a few MeV., which diverge from points near the centres of disintegration. The changes in direction of motion of the electron, due to scattering, can be clearly distinguished. In some cases at least, these electrons are produced by the β -decay of unstable nuclear fragments of short-range.

Observer: Mrs. B. MOORE.



Star of type O_n , which is accompanied by the emission of a beryllium nucleus of energy 150 MeV., and an electron. The Be nucleus, which has been identified by observations on the δ -rays, produces no decay electron at the end of its range. It was therefore probably of mass 10, 11 or 12 mass-units. The deviations in the track of the electron can be clearly distinguished.

Observer: Mrs. J. Cowie.



Star of type O_n accompanied by the emission of a π^- particle of short-range which produces a second disintegration. Several of the tracks of heavy nuclear fragments show numerous δ -rays, and are probably of charge 3 or greater. Observer: Miss B. HULBERT.



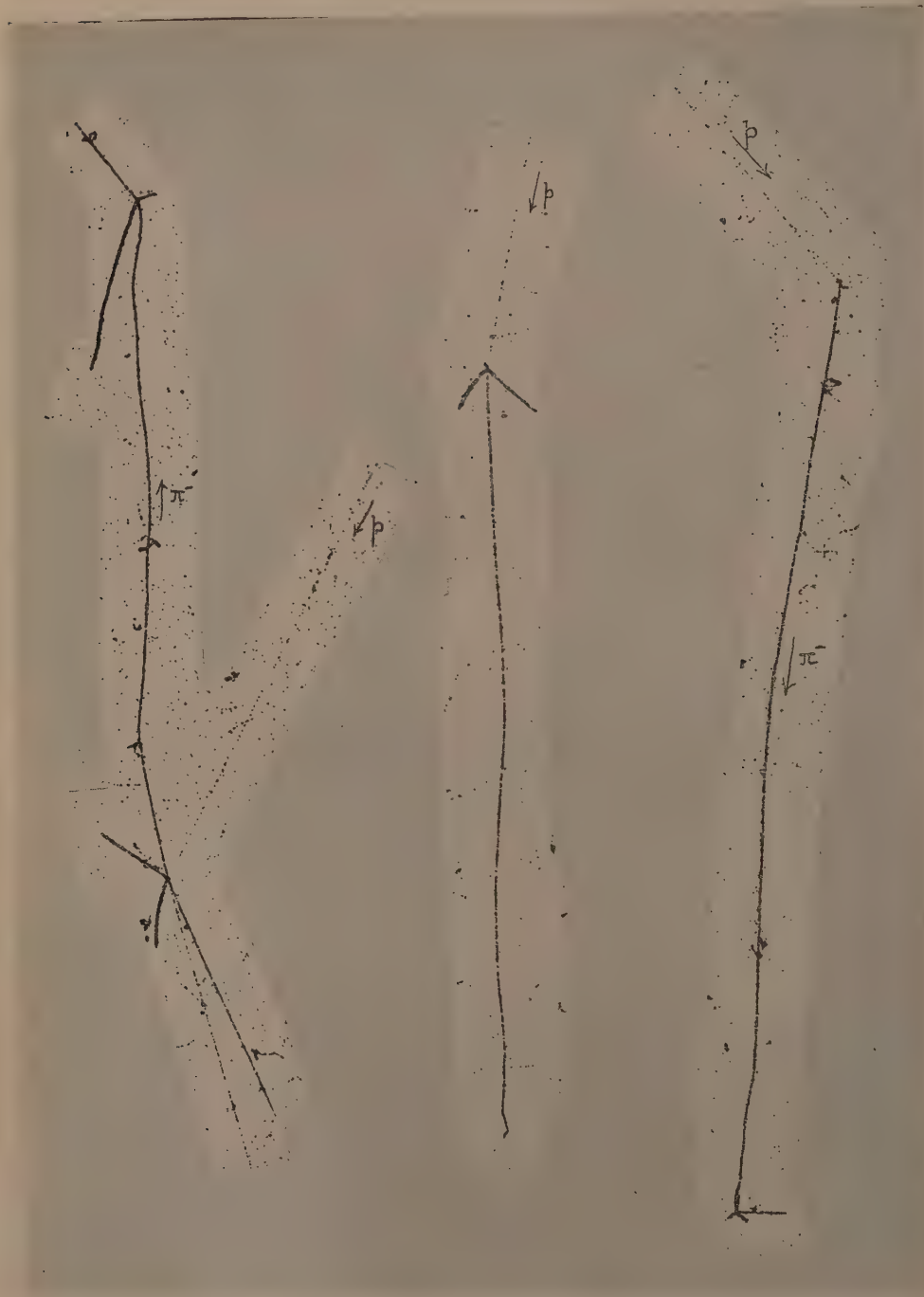
Star of type 6_n ($N_h=4$).

Observer : Miss G. HUSSEY.



Star of type 6_p ($\mathcal{N}_h=4$).

Observer: Miss A. COLE.



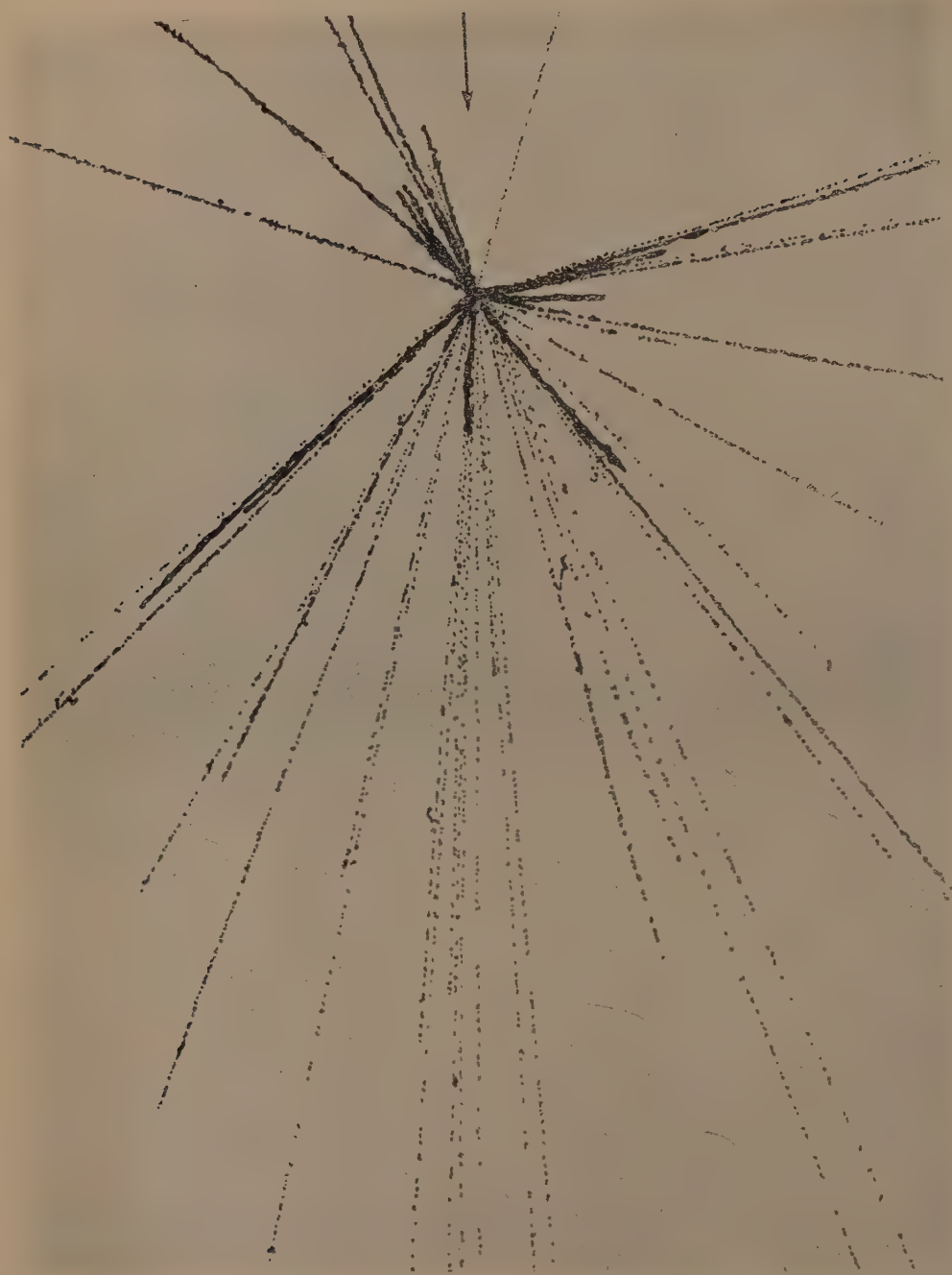
Three examples of "stars" of type O_p which are accompanied by the emission of mesons. In two cases the mesons produce secondary disintegrations. The evidence is consistent with the view that the slow mesons produced in processes of this type are invariably π^- -particles.

Observers : Mrs. J. COWIE and Mrs. I. POWELL.



Mosaic of photo-micrographs of a "shower" of type 10_n ($\mathcal{N}_h=9$).

Observer: Mrs. I. POWELL.

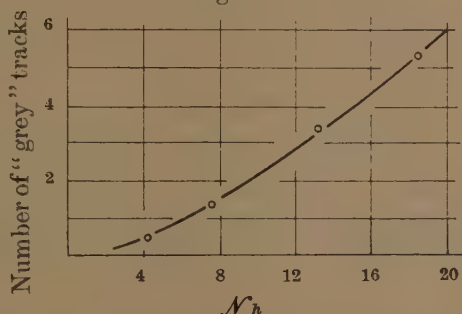


Facsimile drawing, made with the aid of the projection microscope, of "shower" of type 20_n ($N_h=23$).

Observer : Mrs. M. VAN DER MERWE.

in fig. 14. From these observations we can determine, for each class, the mean energy of the particles producing the "grey" tracks. We can also determine the average number of "grey" tracks for stars with different

Fig. 15.



Average number of "grey" tracks as a function of total number of heavily ionizing particles, N_h , in a "star".

values of N_h , see fig. 15. Combining the results, the average total energy per "star" represented by the "grey" tracks, for the different values of N_h , can be found. We assume, once again, that the energy carried by the neutrons is 1.25 as great as that of the protons.

The results are summarized in Table IV., and plotted in fig. 2.

TABLE IV.

Class of star	N_h (mean)	Number per star	Correction factor for neutrons	Energy per particle in MeV.		Total in MeV.
				Kinetic	Binding	
$N_h=4, 5$	4.1	(a) 1.3	1	10	5	20
		(b) 2.3	2.25	10	9	99
		(c) 0.5	2.25	73	9	92
						Total 211 MeV.
$N_h=7, 8$	7.45	(a) 1.4	1	10	5	21
		(b) 4.7	2.25	10	9	201
		(c) 1.35	2.25	85	9	285
						Total 507 MeV.
$N_h=12-15$	13.1	(a) 2.0	1	10	5	28
		(b) 7.9	2.25	10	9	338
		(c) 3.2	2.25	105	9	817
						Total 1183 MeV.
$N_h=16-24$	18.3	(a) 2.7	1	10	5	40
		(b) 10.3	2.25	10	9	440
		(c) 5.3	2.25	121	9	1555
						Total 2035 MeV.

REFERENCES.

- BROWN, CAMERINI, FOWLER, KING, MUIRHEAD, POWELL and RITSON. 1949, *Nature*, **163**, 47 and 82.
 DILWORTH, PAYNE and OCCHIALINI, 1948, *Nature*, **161**, 120.

LXXX. *Notices of New Books and Periodicals received.*

The Unitary Principle in Physics and Biology. By L. L. WHYTE. [Pp. ix+181.]
(The Cresset Press, 1949.) Price 12s. 6d. net.

It is not possible to deal adequately with this book in a short review. In an attempt to find some all-embracing ("unitary") principle which will unite the special sciences, the author suggests that it should take the form: change proceeds so as to reduce the asymmetry in an isolable system.

The principle is a development of, and resembles in its wide generality, the principle of increase of entropy of the universe. In its present form, the present reviewer thinks the principle not sufficiently well-defined to be useful in physics.

Leitfähigkeit und Leitungsmechanismus fester Stoffe. EDUARD JUSTI. [Pp. 339.]
(Göttingen: Venderhoeck and Ruprecht, 1948.)

PROFESSOR JUSTI's book is a very useful source of information on a number of subjects, including thermoelectricity, contacts and rectifiers, semi-conductors and super-conductors, photo-conductors and so on. Particularly interesting are the sections devoted to the change of resistance in a magnetic field, and super-conducting alloys. There are over 200 diagrams and a very large number of references.

The Fundamentals of Electromagnetism. By E. G. CULLWICK. [Pp. xxvi+327.]
(Second edition: Cambridge University Press, 1949.) Price 18s. net.

THIS is a corrected reprint, rather than a new edition, of Professor Cullwick's book, which was first issued in 1939. The standpoint adopted is that of operationalism, and the concept of the unit magnetic pole is resolutely avoided. There is a lucid discussion of the frequently confusing question of units and their interconversion. The theoretical development is based on the classical experiments, interpreted in terms of an interaction between moving charges which depends on their positions, velocities and accelerations. It is made clear, however, that the author is concerned with the interactions of circuits rather than of charges, and to this extent the treatment avoids the real fundamentals.

The book is intended for students of electrical engineering, and many of the well-known paradoxes of electromagnetism are discussed. It should prove stimulating reading, in addition to the usual texts, for students in the Physics Honours school, and for their teachers. G. W.

LXXXI. *Stresses and Strains in Tube-Drawing.*

By H. W. SWIFT *.

[Received June 15, 1949.]

ATTEMPTS to develop a theory of deep-drawing operations have hitherto failed to reach a satisfactory agreement with experiment and practice in either of the fields of drawing from a flat blank, re-drawing from a cylindrical cup or "empty sinking" of a tube. Previous treatments have, in predicting stresses and drawing forces, neglected changes in wall thickness which occur during the process of drawing and have either neglected friction and work-hardening or treated them in a superficial and empirical way.

Basically the problem of tube-drawing, which is similar to that of re-drawing from a cylindrical cup of uniform wall thickness, is simpler than that of deep-drawing from a flat blank, because the boundary conditions at the outer and inner radii are constant during tube-drawing, whereas they change continually during the operation of drawing from a blank. In the present paper an attempt is made to examine in some detail the problem of "empty sinking", *i. e.* drawing without a restrictive mandrel, in order to test the validity of certain simplifying assumptions, to ascertain the probable effects of various impressed conditions, and to compare the results with such relevant experimental data as are available. In this treatment it will be assumed that the ratio t_0/d_0 of the wall thickness to the tube diameter is so small that bending effects can be neglected and the stress assumed uniform through the walls.

Two modes of sinking will be analysed; one in which the motive force is transmitted as tension in the drawn tube, as indicated in fig. 1 (*a*), and the other in which the tube is pressed through, as shown in fig. 1 (*b*). In industry these modes are referred to as "empty sinking" and "pressure sinking" respectively, but they will here be called briefly "drawing" and "sinking". The former of these modes has been studied with certain simplifying assumptions by Sachs and Baldwin (1946), who have obtained expressions for the stress distribution without regard to thickness changes.

Consider part of an elementary ring in a tube at any stage of the reducing operation. Provided that friction between the die and the tube is neglected, symmetry requires that the principal stresses shall be those shown in fig. 2 as p_1 defined as the radial stress, p_2 as the lateral stress, and p_3 as the hoop stress. For analytical consistency these and other normal stresses will be regarded as positive when tensile.

* Communicated by the Author.

Fig. 1.

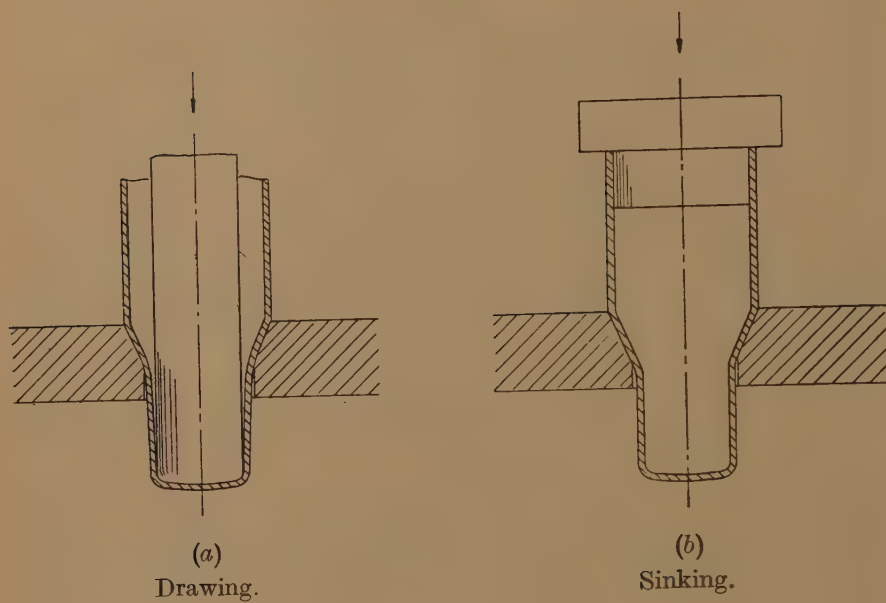
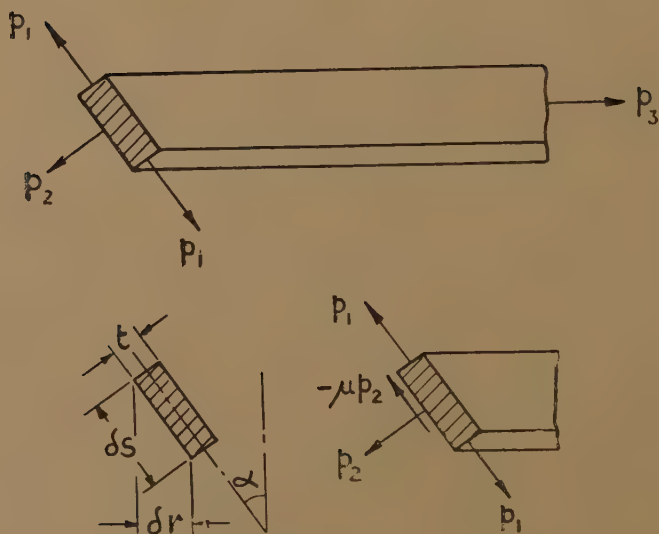


Fig. 2.



With the notation of fig. 2 it is easily shown, by resolving horizontally the forces acting on half the elementary ring, that

$$p_3 t = p_2 r \cos \alpha + \sin \alpha d(p_1 t r \sin \alpha) / dr,$$

and, by resolving vertically for the whole ring,

$$p_2 r = d(p_1 t r \cos \alpha) / dr.$$

Eliminating p_2 between these equations gives

$$p_3 t = d(p_1 t r) / dr = r d(p_1 t) / dr + p_1 t. \quad \dots \dots \dots (i)$$

In tube-drawing, once steady conditions have been attained, all the stresses and strains are functions of the radius r and the function of equation (i) is to assist in the evaluation of p_1 , p_3 and t in terms of r . For this purpose, two further equations connecting these quantities are needed and they must be sought from the conditions of strain. One is provided by the condition of plasticity, *i. e.* the stress relationship under which plastic flow occurs, while the other is provided by the relationship between the strains which constitute this flow, and, in particular, by the conditions controlling changes in the thickness t .

Two hypotheses regarding the stress conditions for plastic flow are widely held at the present time; one adopts the maximum shear-stress as the criterion and the other the distortional resilience. In cases where strain in one principal direction is suppressed, as would arise, for example, if the wall thickness of a tube were kept constant during drawing, the two hypotheses lead to exactly similar analysis and their results vary only by a constant factor $2/\sqrt{3}$ between corresponding stresses. In the case of "empty" drawing, however, this condition is not fulfilled and it will be well to determine to what extent the results of the two hypotheses are at variance. This matter is of some importance because the shear-stress hypothesis is considerably easier to handle analytically, particularly when thickness changes are taken into account.

In the first place, therefore, we shall make a preliminary comparison between the results given by the two hypotheses in drawing and sinking, when thickness changes are neglected in evaluating the stresses, but subsequently estimated in the light of these stresses.

COMPARISON OF YIELD HYPOTHESES.

If changes in thickness are neglected, equation (i) becomes

$$p_3 = p_1 + r dp_1 / dr.$$

Adopting first the shear resilience hypothesis, this gives the same form of stress relationship whether the tube is drawn or sunk

$$(p_2 - p_3)^2 + (p_3 - p_1)^2 + (p_1 - p_2)^2 = 2f^2,$$

where f is the yield stress in simple tension. Of the three principal stresses, p_2 is always relatively small in a thin-walled tube—if the die-angle α is constant, $p_2 = (p_3 t \cos \alpha) / r$ —and for all ordinary purposes may be neglected. In this case the plasticity condition can be written in the form

$$p_3 = (p_1 - \sqrt{(4f^2 - 3p_1^2)}) / 2, \quad \dots \dots \dots (ii)$$

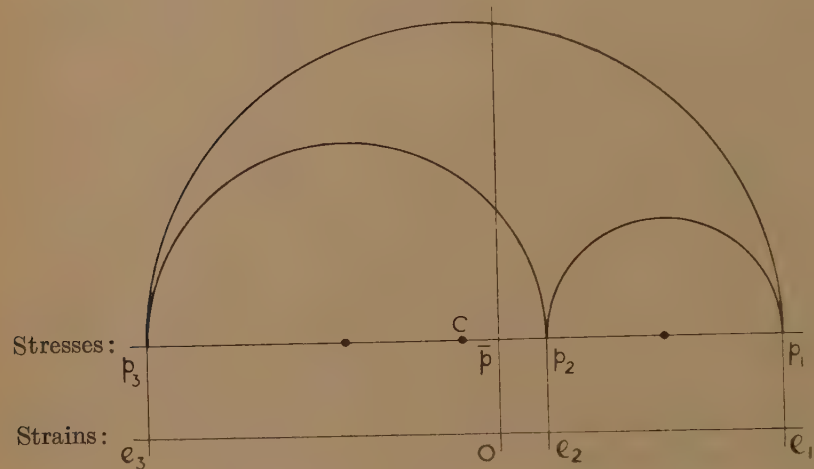
and equation (i) can then be written

$$dr/r + 2\,dx/(x + \sqrt{4 - 3x^2}) = 0, \quad \text{where } x = p_1/f.$$

This, of course, provides on integration the relationship between the radial stress p_1 and the current value of the radius r , and from this the distribution of p_3 can be found from the plasticity condition (ii).

The thickness changes which, although neglected in these stress calculations, are demanded by them, can be obtained from a knowledge of the stress-strain relationships for plastic flow. Although it is now known to be subject to correction with the strain-hardening metals commonly used in drawing operations, the Levy-Lode relationship is the simplest working approximation. This postulates the similarity of the Mohr stress and strain increment circles as indicated in fig. 3.

Fig. 3.



In the notation of the figure, e_2 represents the proportional increment of thickness strain $\delta t/t$ during a corresponding change in radius $\delta r/r = e_3$. Hence

$$(\delta t/t)/(\delta r/r) = e_2/e_3 = (p_2 - \bar{p})/(p_3 - \bar{p}) = (2p_2 - p_1 - p_3)/(2p_3 - p_1 - p_2).$$

Since this relationship will be used under various conditions later in this paper, it should be noted that its truth is independent of the relative values, or signs, of the principal stresses p_1, p_2, p_3 .

It has been pointed out that, for present purposes, the stress p_2 is negligible by comparison with p_1 and p_3 . Hence, we may write

$$(\delta t/t)/(\delta r/r) = (p_1 + p_3)/(p_1 - 2p_3) \quad \dots \dots \dots (iii)$$

and, by virtue of the constancy of volume,

$$\delta t/t : \delta r/r : \delta l/l = (p_1 + p_3) : (p_1 - 2p_3) : (-2p_1 + p_3),$$

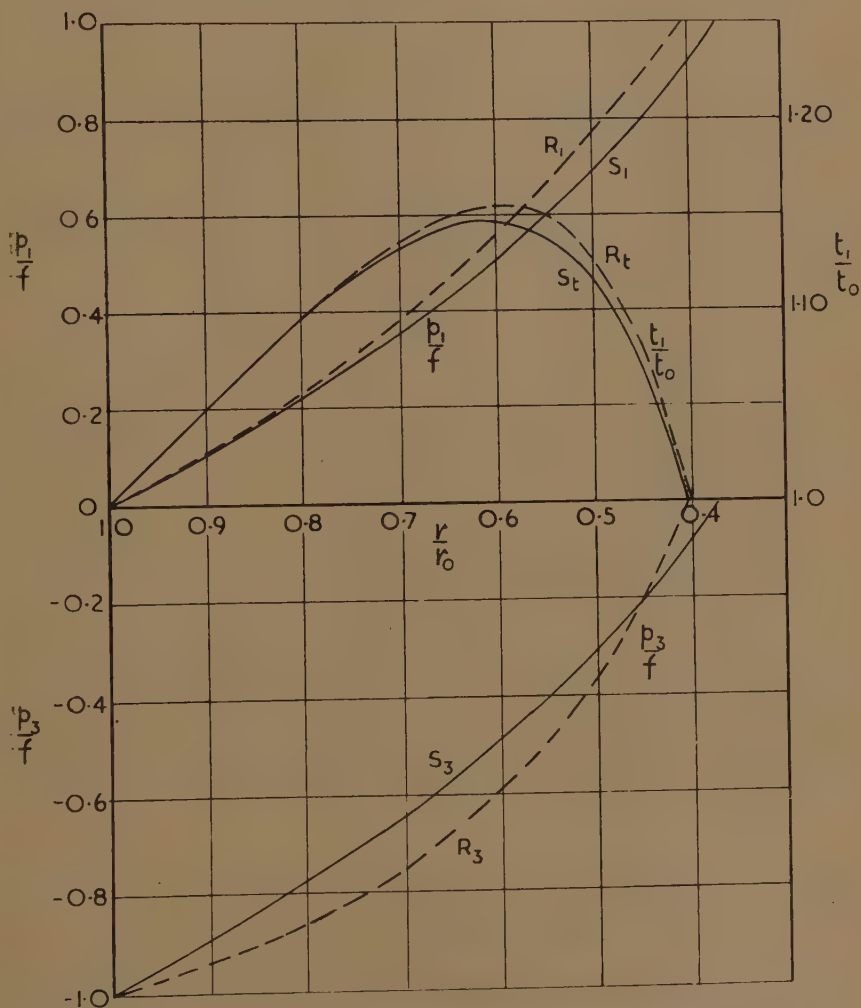
where $\delta l/l$ is the axial strain increment in the direction of p_1 .

If the tube is drawn from an initial radius r_0 to a smaller radius r_1 , the drawing stress and thickening strain at any radius r are given by

$$\ln (r/r_0) + (\ln \sin (\theta + \pi/3) - \ln \sin (\pi/3))/2 + \sqrt{3}\theta/2 = 0, \quad (\text{iv})$$

$$t/t_0 = 2 \cos (\theta - \pi/6)/\sqrt{3}, \quad \text{where} \quad \sin \theta = \sqrt{3}p_1/2f. \quad (\text{v})$$

Fig. 4.



Drawing:—Comparison of yield hypotheses.

S, Shear-stress hypothesis. R, Shear-resilience hypothesis.

The stress ratios p_1/f , p_3/f and the thickening strain $\Delta t/t_0$ for various values of r/r_0 are plotted in fig. 4 by means of curves marked R_1 , R_3 and R_t respectively.

If we now apply the shear-stress hypothesis to the problem of drawing, the appropriate plasticity condition is $p_1 - p_3 = f$, and the solution of equation (i) becomes simply

$$p_1/f = \ln(r_0/r), \quad p_3/f = \ln(r_0/r) - 1,$$

while thickness changes are given by substitution in and integration of equation (iii)

$$\ln(t/t_0) = 2x + 3 \ln(1 - x/2), \quad \text{where} \quad x = p_1/f = \ln(r_0/r).$$

The stress ratios and thickening strain according to the shear-stress hypothesis are plotted in fig. 4 as curves marked S_1 , S_3 , S_t respectively.

It will be seen that so far as drawing is concerned the two hypotheses lead to very similar values of the drawing stress p_1 and of the thickness changes over the whole of the useful range. In practice the reduction obtained in a single stage seldom exceeds 30 per cent. The hoop stresses predicted by the two hypotheses are more at variance, but, since these stresses are compressive, they do not limit the drawability and are therefore only of secondary importance in practice.

Turning now to compare the two hypotheses under conditions of sinking, the shear resilience hypothesis leads to the same equations (iv) as in drawing, provided that the inner (stress-free) radius r_1 be substituted for r_0 and negative values given to the stress ratio $x = p_1/f$. For purposes of comparison with drawing conditions, it is desirable to express the stresses and strains in terms of reduction from r_0 .

In this case the values at intermediate values of r are not single-valued functions of r/r_0 , as in drawing, but depend on r_1/r_0 as well. Nevertheless the stresses at r_0 required to sink the tube from r_0 to r_1 and the resultant thickening strain induced by this operation may all be plotted as single curves in terms of r_1/r_0 , and these are shown as R_1 , R_3 and R_t in fig. 5.

When the shear-stress hypothesis is applied to sinking, since the radial stress p_1 is always compressive, the maximum shear-stress is $(p_2 - p_3)/2$ and the appropriate plasticity condition is $p_2 - p_3 = f$, or, since p_2 is negligible, $p_3 = -f$. With this substitution it is easily shown from equations (i) and (iii) that

$$p_1/f = r_1/r - 1,$$

$$\ln(t/t_1) = \ln(r/r_1) - 3 \ln(r + r_1)/2r_1,$$

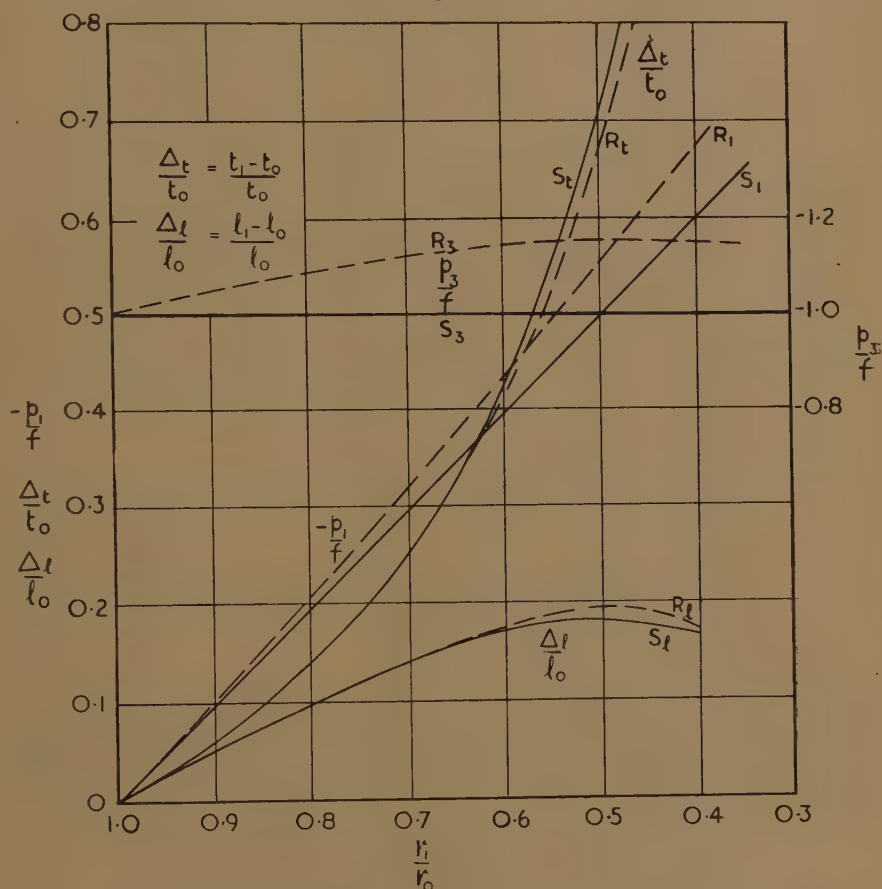
$$t_0/t_1 = r_1(r_0/r_1 + 1)^3/8r_0.$$

When plotted in terms of the initial radius r_0 , these expressions lead to the curves shown as S_1 , S_3 and S_t in fig. 5.

A comparison of curves R , S in this figure shows that, although there is a greater discrepancy between the sinking stresses p_1 than between the drawing stresses in fig. 4, the predicted thickness changes are almost identical over the whole range. Moreover, the divergence between sinking-stress values over the practical range is insufficient to necessitate separate examination of the more general problems in terms of both hypotheses.

Having regard to the results obtained above with drawing and sinking, it seems clear that theoretical work based on either of the two hypotheses will lead to very similar results (particularly in regard to strains) over the practical range, and that the choice between the two criteria can be made on grounds of convenience. The shear-stress criterion, being the simpler, will therefore be adopted.

Fig. 5.



Sinking:—Comparison of yield hypotheses.

S, Shear-stress. R, Shear-resilience.

FRICTIONLESS AND NON-HARDENING MATERIAL.

Adopting the shear-stress criterion, we shall now discuss the drawing or sinking of a tube taking proper account of changes in thickness, but for the moment neglecting friction and strain-hardening.

Under drawing conditions the plasticity condition $p_1 - p_3 = f$ when substituted in equation (i) gives

$$\delta p_1 + p_1 \delta t/t + f \delta r/r = 0,$$

while the stress-strain correlation takes the form

$$\delta t/t = \delta r/r \times (2p_1 - f)/(2f - p_1),$$

and constancy of volume requires

$$\delta r/r + \delta t/t + \delta l/l = 0.$$

Integration gives

$$\ln(r/r_0) = \frac{1}{4} \ln(1 - x + x^2) - (\sqrt{3}/2) \tan^{-1}(x\sqrt{3}/2 - x),$$

$$\ln(t/t_0) = -\frac{1}{2} \ln(1 - x + x^2),$$

$$\ln(l/l_0) = \frac{1}{4} \ln(1 - x + x^2) + (\sqrt{3}/2) \tan^{-1}(x\sqrt{3}/2 - x), \quad . \quad (v)$$

where $x = p_1/f$ and is a parameter of correlation between r , t , and l . Values of x appropriate to drawing are essentially positive.

Curves showing p_1/f , $\Delta t/t_0$ and $\Delta l/l_0$ in terms of r/r_0 are shown as D_1 , D_t , D_l in fig. 6. These curves show the stress and strain distribution for any drawing ratio, the curve for a particular ratio r_1/r_0 simply terminating at the ordinate corresponding to this ratio.

Under sinking conditions the plasticity condition $p_3 = -f$ applied to equation (i) gives

$$\delta p_1 + p_1 \delta t/t + (f + p_1) \delta r/r = 0,$$

and the stress-strain correlation takes the form

$$\delta t/t = \delta r(p_1 - f)/r(p_1 + 2f).$$

$$\ln(r/r_1) = -\frac{1}{4} \ln(1 + z + z^2) - (\sqrt{3}/2) \tan^{-1}(z\sqrt{3}/2 + z),$$

$$\ln(t/t_1) = -\frac{1}{4} \ln(1 + z + z^2) + (\sqrt{3}/2) \tan^{-1}(z\sqrt{3}/2 + z),$$

$$\ln(l/l_1) = \frac{1}{2} \ln(1 + z + z^2), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (vii)$$

where $z = p_1/f$ which in this case, by contrast with drawing, has essentially negative values.

If for the moment we consider the resultant strains and final (applied) stress for a reduction from r_0 to r_1 , then the two cases give equations of the type

$$\ln(r_1/r_0) = \frac{1}{4} \ln(1 - x_1 + x_1^2) - (\sqrt{3}/2) \tan^{-1}(x\sqrt{3}/2 - x_1),$$

$$\ln(r_0/r_1) = -\frac{1}{4} \ln(1 + z_0 + z_0^2) - (\sqrt{3}/2) \tan^{-1}(z_0\sqrt{3}/2 - z_0).$$

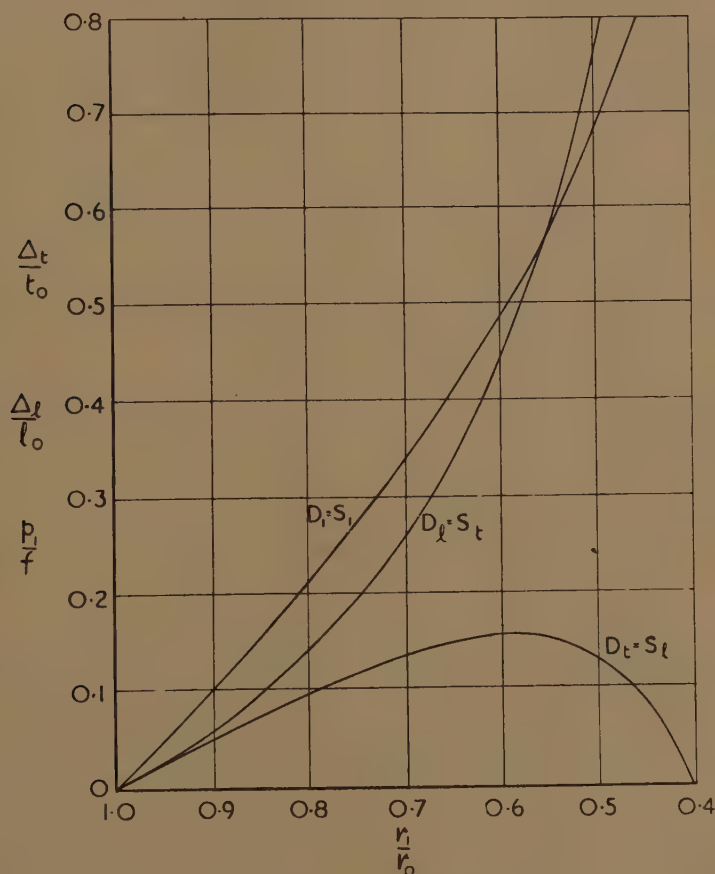
Hence clearly for a given reduction ratio r_1/r_0 , $z_0 = -x_1$, and the tensile drawing stress required at radius r_1 is identical in value with the compressive sinking stress required at radius r_0 . In a similar way it will be seen that the resultant thickening strain $t_1/t_0 - 1$ in a drawing operation is identical with the resultant longitudinal strain ratio $l/l_0 - 1$ in the corresponding sinking operation, and *vice versa*. These analogies are made clear in fig. 6 by comparison of curves $D_1 = S_1$, $D_t = S_t$, $D_l = S_l$ respectively.

It will be clear that either of the curves of l_1/l_0 , t_1/t_0 can be regarded as consequential on the other for a given reduction r_1/r_0 , and therefore it is sufficient that one or the other should be used as a basis of comparison

between different conditions. Because the forms of the t_1/t_0 curves in drawing and the l_1/l_0 curves in sinking are more compact and discriminating than their counterparts, these will in fact be used in comparisons later in this paper.

Since the drawing stresses and strains are all expressed in terms of the initial values r_0 , t_0 , l_0 , their values at any radius r are independent of the reduced radius r_1 and the curves D_1 , D_t , D_l can be used to measure either

Fig. 6.



Drawing and sinking:—No friction, no hardening.

the applied drawing stress and the resultant strains in a complete drawing operation or the current values of these quantities during the course of such an operation. On the other hand, the corresponding expressions for sinking conditions are in terms of the final radius r_1 (at which the radial stress is zero) and can therefore only be used, as they stand, to obtain the applied sinking stress and the resultant strains.

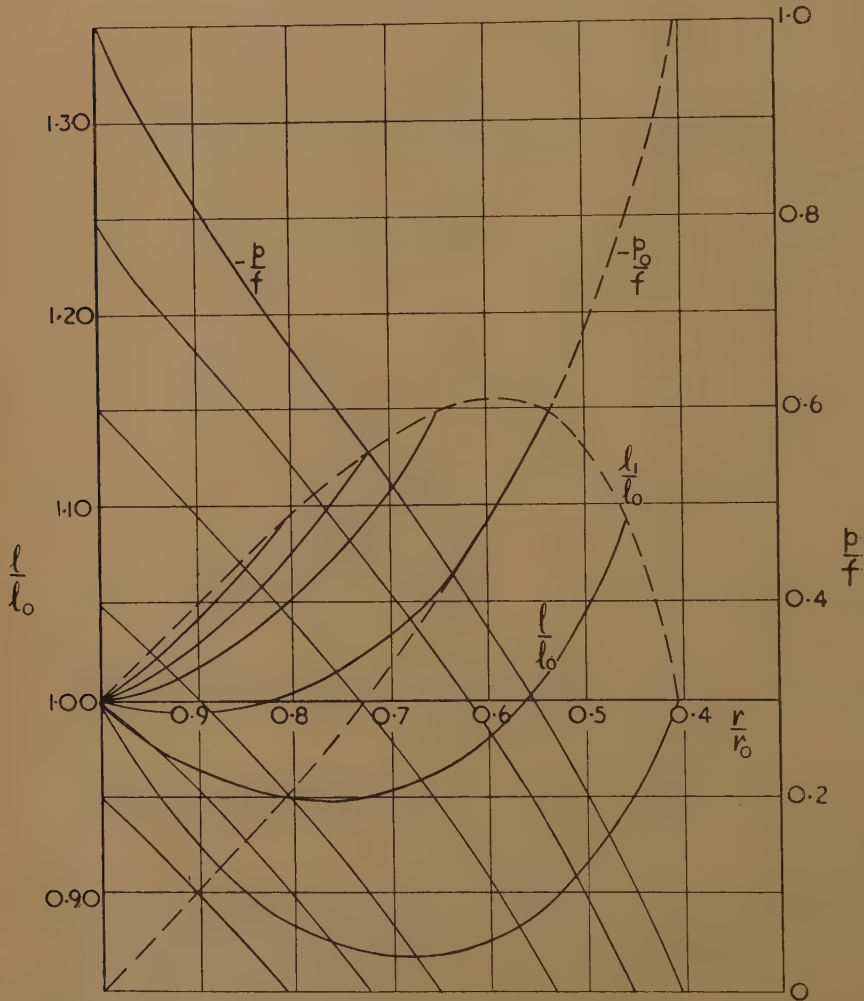
In order to ascertain the distribution or current values of the strains during sinking, it is necessary to make use of derived formulæ of the type

$$t/t_0 = (t/t_1)(t_1/t_0),$$

$$l/l_0 = (l/l_1)(l_1/l_0),$$

in which the ratios t_1/t_0 and l_1/l_0 are obtained from the basic equations (vii),

Fig. 7.



by correlation with the value of z_0 appropriate to the known reduction ratio r_1/r_0 . The distribution or current values of radial stress can of course be obtained directly from the relation

$$\ln(r/r_0) = -\frac{1}{2} \ln(1+z+z^2) - (\sqrt{3}/2) \tan^{-1}(z\sqrt{3/2}+z) + \ln(r_1/r_0).$$

The distribution of radial stresses and axial strains in sinking operations of different severity, from a given radius r_0 are plotted in fig. 7, in contrast with the applied stress p_0 and resultant strain $(l_0/l_1 - 1)$ for these operations, which are shown by broken lines.

INFLUENCE OF FRICTION AND STRAIN-HARDENING.

Still adopting shear-stress as the criterion of plastic yield, we shall now consider the problems of drawing and sinking, taking account of the effects of strain-hardening and die friction as well as changes in thickness.

When friction between the die and tube is taken into account, the stress system is modified generally as indicated in fig. 2 (*a*). The disposition shown is of course theoretically incomplete, because p_2 is no longer a principal stress, and p_1 on the perpendicular plane must therefore be associated with a (non-uniform) shear-stress. But this has no significant effect on statical conditions, and by resolving forces horizontally and vertically we find

$$p_3 t = (1 - \mu \tan \alpha) p_2 r \cos \alpha + \sin \alpha d(p_1 t r \sin \alpha) / dr,$$

$$p_2 r (1 + \mu \cot \alpha) = d(p_1 t r \cos \alpha) / dr.$$

If the die angle is constant, elimination of p_2 gives

$$p_3 t (1 + \mu \cot \alpha) = d(p_1 t r) / dr. \quad \dots \dots \dots (viii)$$

The two further equations necessary to determine p_1 , p_3 , t in terms of r are obtained from the conditions of constancy of volume and of plastic yield. Constancy of volume again requires $\delta r/r + \delta l/l + \delta t/t = 0$ and the shear-stress criterion for plastic yield is also substantially unaffected by friction :

$$\begin{aligned} &\text{for drawing } p_1 - p_3 = f, \\ &\text{for sinking } -p_3 = f, \end{aligned}$$

where f is the tensile yield stress for the material at any radius r .

This stress cannot now be assumed as constant but will depend on the initial yield stress f_0 and on the degree of strain-hardening induced during the drawing of the material from radius r_0 to radius r . In order to handle the problem analytically, it is necessary to relate the stress f in some simple way with the other operative conditions, preferably with the stresses p_1 , p_3 . In drawing, the material enters with yield stress f_0 and with radial stress $p_1 = 0$. During the drawing process this radial stress increases continuously and, when friction and hardening are neglected, is proportional to the logarithmic principal strain

$$p_1 = f_0 \ln (r_0/r).$$

If, therefore, we assume a uniform rate of strain-hardening in relation to this strain, we may write $f = f_0 + c p_1$, where c is a measure of the strain-hardening rate. By analogy, in the case of sinking, it seems reasonable to assume a relation of the form $f = f_0 + c(p_1 - p_0)$, where p_0 is the value of p_1 at entry to the die.

I. *Drawing.*

When the appropriate substitution for f is made in equation (viii), this can be written in the form

$$\frac{dr}{r} = \frac{d(p_1 t)}{p_1[(1-c)(1+m)-1]-f_0(1+m)},$$

where $m = \mu \cot \alpha$. Now

$$\frac{dt}{t} = \frac{dr}{r} \frac{p_1 + p_3}{p_1 - 2p_3} = \frac{dr}{r} \cdot \frac{p_1(2-c)-f_0}{-p_1(1-2c)+2f_0}.$$

Elimination of t between these equations gives

$$\frac{dr}{r} = \frac{[2-x(1-2c)]dx}{-2(1+m)+x(2+3m-4c-4m)-x^2(2+m-2c-3cm+2c^2+2c^2m)},$$

where $x = p_1/f_0$. This expression can of course be integrated by analytical methods, but the general expression is unwieldy and integration in any particular case is more simply carried out after direct substitution in the above equation. By writing $c=0$, for example, we find the effect of friction without strain-hardening, and by writing $m=0$ the effect of strain-hardening without friction.

(a) With friction, but no strain-hardening, the equation becomes

$$dr/r = (2-x) dx/X,$$

where

$$X = -2(1+m) + x(2+3m) - x^2(2+m),$$

and integrates in the form

$$\ln(r/r_0) = (\ln(X/X_0))/(2(2+m) - (6+m)\phi/(2+m)),$$

where

$$\tan \sqrt{(12+12m-m^2)\phi} = x \sqrt{(12+12m-m^2)/[4m+4-x(2+3m)]}.$$

This, of course, gives the distribution and changes of drawing stress p_1 during the drawing process.

The thickness changes and third principal strain are given by similar expressions

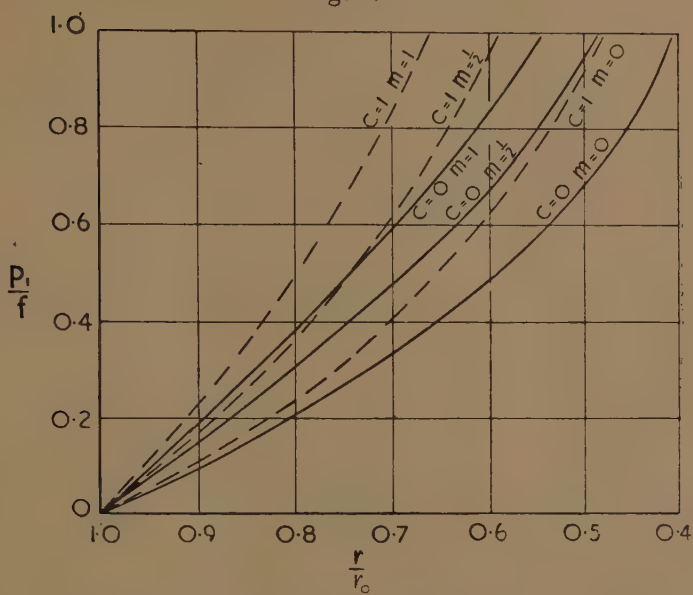
$$\ln(t/t_0) = -(\ln(X/X_0))/(2(2+m) - 4m\phi/(2+m)),$$

$$\ln(l/l_0) = (\ln(X/X_0))/(2(2+m) + (6+5m)\phi/(2+m)).$$

The stress and strain values for certain reasonable values of m are plotted in figs. 8, 9. With a die angle $\alpha = 15^\circ$, for example, $m = \frac{1}{2}$ corresponds to a coefficient of friction $\mu = 0.134$ and $m = 1$ to $\mu = 0.27$. It will be seen from fig. 8 that this latter degree of friction increases the drawing stress (and therefore the drawing load) by some 60 per cent. over the normal range of reductions.

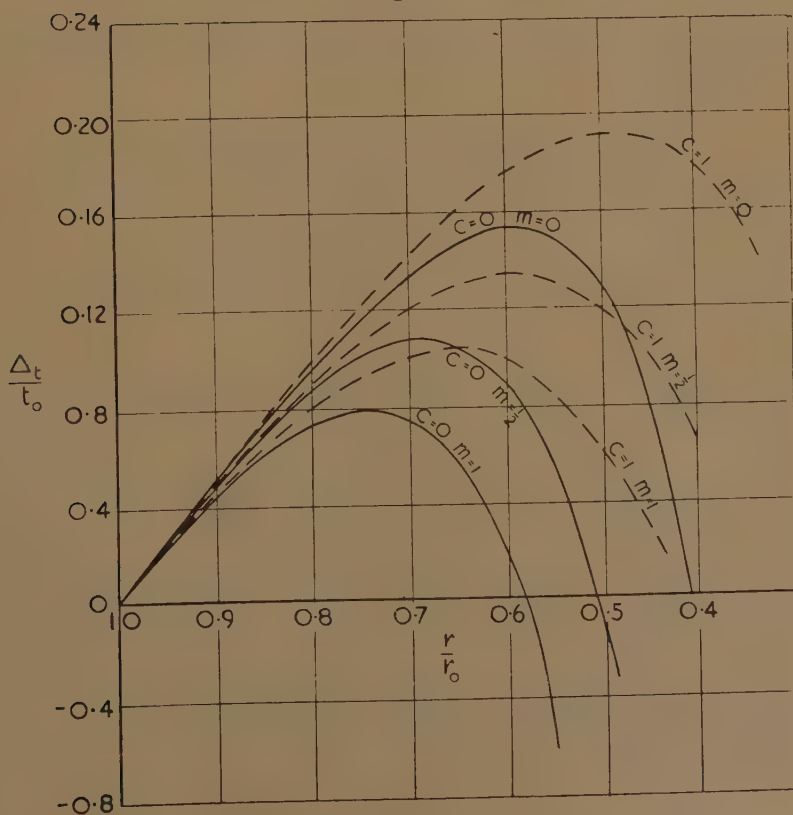
At the same time the thickening of the metal is substantially reduced and the length of the tube correspondingly increased.

Fig. 8.



Drawing stress.

Fig. 9.



(b) With strain-hardening, but no friction, the solutions take the form

$$\ln (r/r_0) = A(1-2c)/4(1-c+c^2) - \psi\sqrt{3}/2(1-c+c^2),$$

$$\ln (t/t_0) = -A(2-c)/4(1-c+c^2) + \psi\sqrt{3}/2(1-c+c^2),$$

$$\therefore \ln (l/l_0) = A(1+c)/4(1-c+c^2) + \psi(1-c\sqrt{3})/2(1-c+c^2),$$

where

$$A = \ln [(1-c+c^2)x - (1-2c)x + 1],$$

$$\tan \psi = x\sqrt{3}/[2-x(1-2c)].$$

The corresponding stresses and strains are plotted in figs. 8, 9 for certain values of c . It will be seen that the effect of strain-hardening in a drawing operation is to increase the drawing stresses and load, as would be expected, and to increase the metal thickness, but only to a small extent within the practical drawing range.

In order to show the effect of friction in combination with strain-hardening, calculations have been made for the two combinations $c=1$, $m=\frac{1}{2}$ and $c=1$, $m=1$, and the results are plotted in figs. 8, 9. It will be seen that the two factors conspire together to increase the drawing stress; $c=1$, $m=1$ gives stresses $2\frac{1}{2}$ times those required with no friction or strain-hardening. On the other hand, the two effects tend to cancel one another in regard to thickness changes and other strains, though the frictional effect appears to predominate in the practical range.

II. *Sinking.*

In this case the appropriate substitution $f=f_0+c(p_1-p_0)$ gives equation (iii) in the form

$$\frac{dr}{r} = \frac{-d(p_1 t)}{p_1[1+c(1+m)] + (f_0 - cp_0)(1+m)} \cdot \frac{1}{t},$$

while

$$dt/t = \frac{dr}{r} \cdot \frac{p_1(1-c) - (f_0 - cp_0)}{p_1(1+2c) + 2(f_0 - cp_0)}.$$

The consequent stress equation is

$$-dr/r = [2+x(1+2c)] dx/Z,$$

where

$$Z = 2(1+m) + x(2+m+4c+4cm) + x^2(2+2c+2c^2+cm+2c^2m).$$

In this case $x=p_1/(f_0-cp_0)$ and its useful values are negative. This equation can be solved in a similar way to that for drawing, but, since the stress p_0 is implicit, the stress p_1 cannot be found in terms of x until p_0 is determined, unless $c=0$. The solution is in general of the form: $\ln (r/r_1) = F(x)$ and gives at entry: $\ln (r_0/r_1) = F(x_0)$, where $x_0 = p_0/(f_0 - cp_0)$ so that $p_0/f_0 = x_0/(1+cx_0)$. Hence, by assuming values of x_0 , it is possible to correlate p_0/f_0 with $\ln (r_0/r_1)$ and so obtain the stresses at entry for a given reduction from r_0 to r_1 . It is to be noted that this procedure does not give directly the stresses—or by derivation the strains—at intermediate points; if required, these can be obtained by re-substitution for p_0 , when evaluated, in the general equation.

(a) With friction, but no strain-hardening, this equation reduces to

$$\frac{dr}{r} + \frac{(x+2) dx}{2x^2 + x(2+m) + 2(1+m)} = 0,$$

where $x = p_1/f_0$, which gives

$$\ln(r/r_1) = -\frac{1}{4} \ln(X'/X'_1) - (6-m)\phi'/2,$$

$$\ln(t/t_1) = -\frac{1}{4} \ln(X'/X'_1) + (6+m)\phi'/2,$$

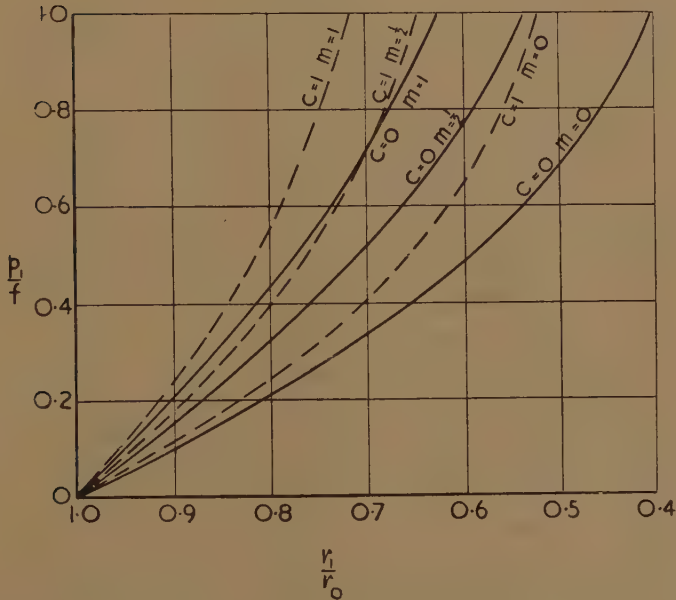
$$\therefore \ln(l/l_1) = \frac{1}{2} \ln(X'/X'_1) - m\phi',$$

where

$$X' = 2x^2 + x(2+m) + 2(1+m); \quad X'_1 = 2(1+m),$$

$$\tan \phi' \sqrt{(12+12m-m^2)} = \frac{x\sqrt{(12+12m-m^2)}}{4(1+m) + x(2+m)}.$$

Fig. 10.



In this case

$$\ln(r/r_0) = -\frac{1}{4} \ln X' - (6-m)\phi'/2 + \frac{1}{4} \ln 2(1+m) + \ln(r_1/r_0),$$

and since r_1/r_0 is known, the stress distribution x can be obtained in terms of r without previously evaluating $x_0 = p_0/f$. But the corresponding relationships for strains t/t_0 , l/l_0 can only be obtained when t_1/t_0 and l_1/l_0 have been found in terms of x_0 from such relations as

$$\ln(t_0/t_1) = -\frac{1}{4} \ln X'_0 + (6+m)\phi'_0/2 + \frac{1}{4} \ln 2(1+m).$$

The leading stresses and strains for certain values of $m = \mu \cot \alpha$ are plotted in figs. 10, 11.

(b) When strain-hardening is present without friction, the general equation becomes

$$\frac{dr}{r} + \frac{dx}{2} \frac{2+x(1+2c)}{x^2(1+c+c^2)+x(1+2c)+1} = 0,$$

where $x = p_1/(f_0 - cp_0)$, which gives

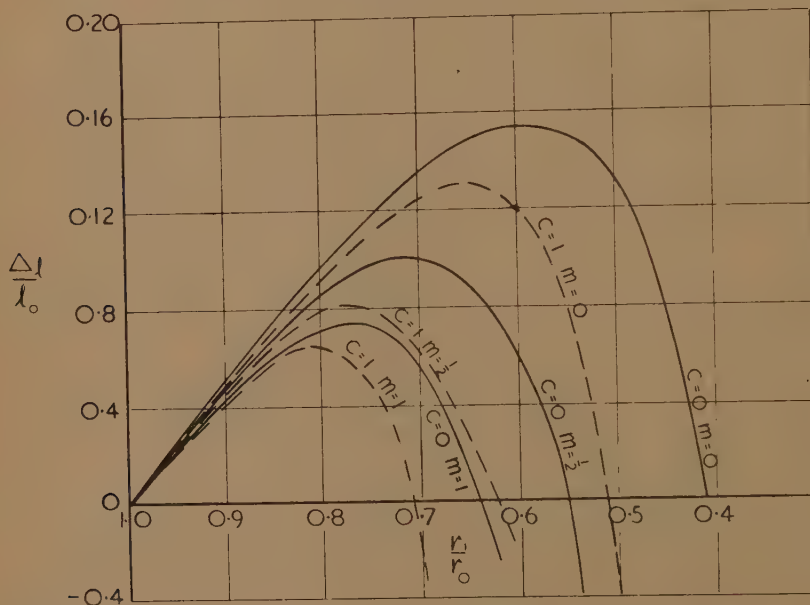
$$\ln\left(\frac{r}{r_1}\right) = -\frac{1+2c}{4(1+c+c^2)} \ln X - \frac{3}{2(1+c+c^2)} \theta,$$

where $\tan \theta = x\sqrt{3}/[2+x(1+2c)]$, $X = x^2 + (2cx + x + 1)/(1+c+c^2)$,

$$\ln(t/t_1) = -\frac{1-c}{4(1+c+c^2)} \ln X + \frac{(1+c)\sqrt{3}\theta}{2(1+c+c^2)}$$

and a consequential expression for $\ln l/l_1$.

Fig. 11.

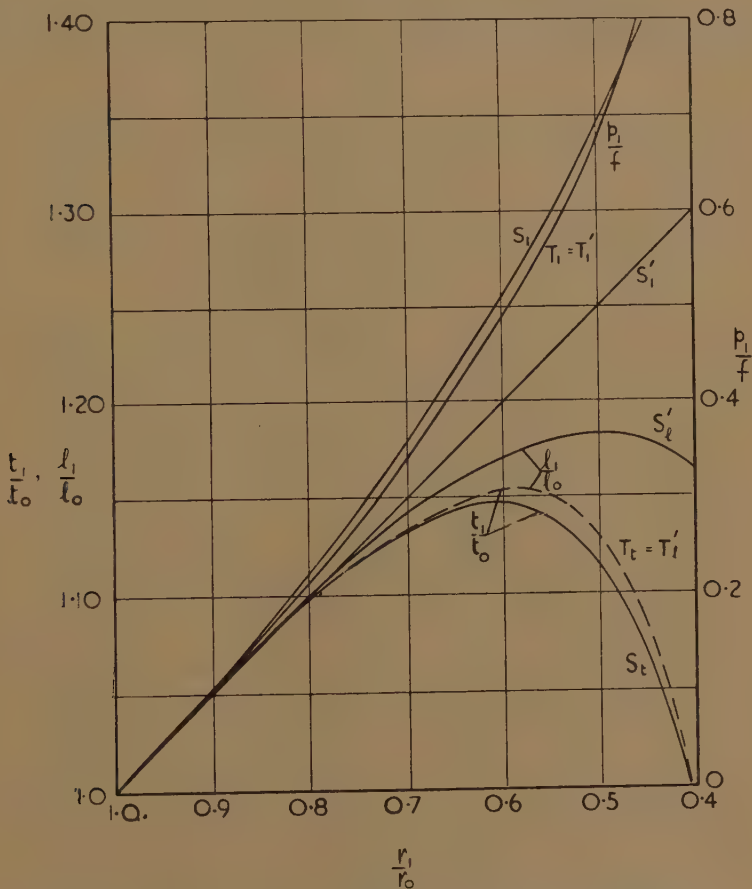


In this case the stress at entry is given by $p_0/f_0 = x_0/(1+cx_0)$ and re-substitution is needed to obtain the stress distribution. Values of the entry stress and of overall thickness and longitudinal strains are shown in figs. 10 and 11, which may be compared with figs. 8 and 9 for comparison with drawing conditions. In making such comparison, it should be noted that the symbol c for a given material will not, in general, have the same value in the two cases, though the difference will not be great over the range of practical drawing.

Calculations have also been made for certain combinations of friction and strain-hardening in the sinking operation and the results are included in figs. 10, 11. The influence of these factors on stress is very similar to that in the drawing operation, but in this case, since both factors tend to increase the thickening and reduce the elongation, their effects on strains are cumulative and not mutually opposed as in drawing.

In comparing the stress ratios p/f_0 for drawing and sinking, it should be noted that, although these determine the press loads required for the operation in both cases, these loads are not directly comparable from the plotted curves of stress. For, whereas the sinking stress is applied across the initial sectional area of the tube, the drawing stress is applied across the final area and should therefore be multiplied by the factor $r_1 t_1 / r_0 t_0 = l_0 / l_1$ for comparison with the sinking stress when press loads are under consideration.

Fig. 12.



Drawing and sinking:—Without friction or hardening.

Radial Tensions: S_1, T_1 drawing. S'_1, T'_1 sinking.

Thickness ratios: S_t, T_t drawing.

Axial strains: S'_l, T'_l sinking.

S, Thickening neglected. T, Thickening considered.

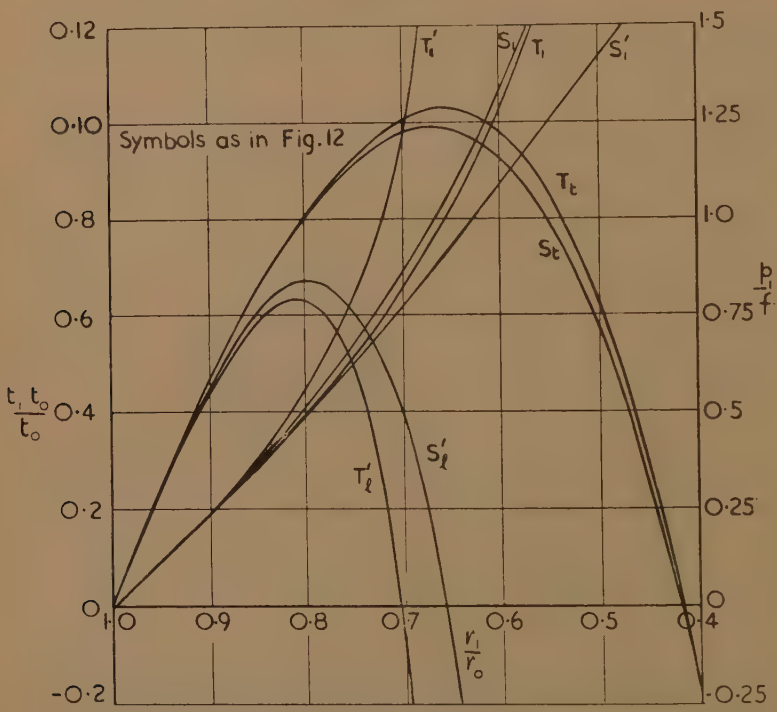
COMPARISON WITH APPROXIMATE THEORY.

It is interesting to compare the results of the theory just developed with those which would have been obtained if changes in thickness were neglected in the stress calculations. Such a comparison is made in figs. 12, 13,

in which curves are plotted showing the applied stress and a representative consequential strain in drawing and sinking, as predicted by the two theories. Curves marked S derive from the more elementary theory, and those marked T, from that which takes thickness changes into account. Fig. 12 refers to a case in which friction and strain-hardening are neglected, while fig. 13 takes both these factors into account, being in fact computed for $c=1, m=1$ in the appropriate equations.

So far as drawing is concerned (S, T curves), it will be seen that in both cases the error involved by neglecting thickness changes in stress calculations is remarkably small over the useful range of reductions ; the stresses and consequential strains predicted by the two theories are almost identical..

Fig. 13.



Drawing and sinking with friction and strain-hardening.

$$\left. \begin{matrix} c=1 \\ m=1 \end{matrix} \right\}.$$

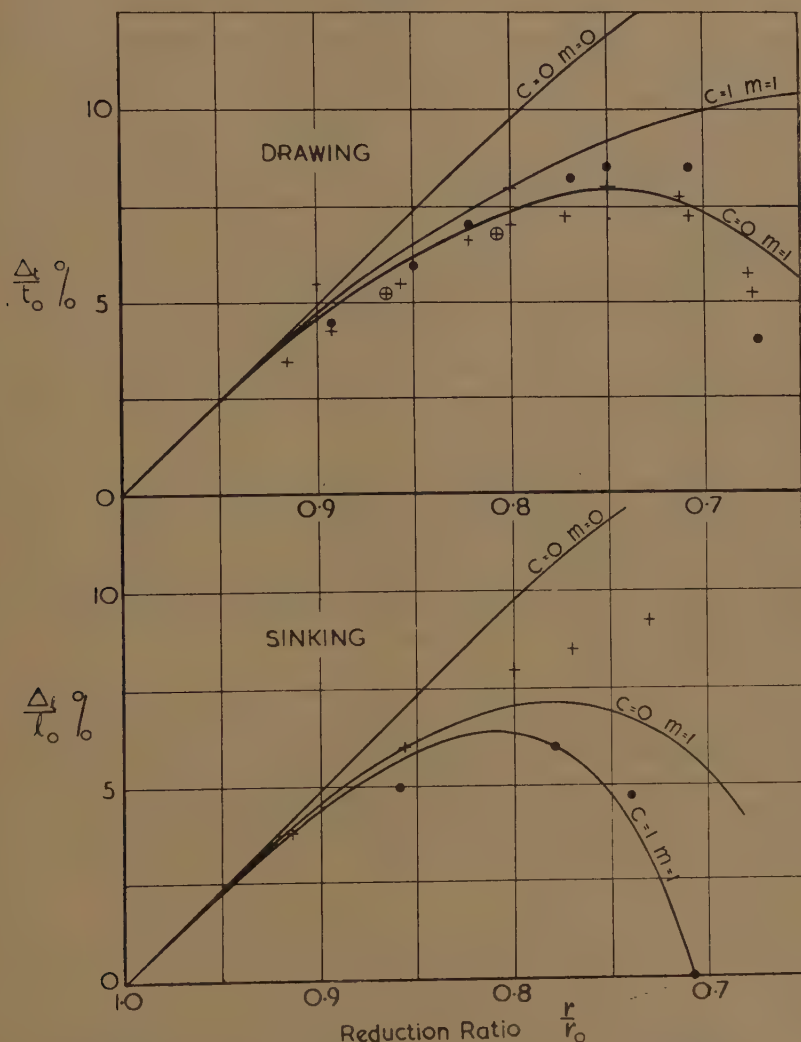
In sinking operations (S', T' curves) the error is not great with low friction and a non-hardening material (fig. 12), but it becomes substantial with more normal materials (fig. 13) when applied to reductions above about 20 per cent.

The similarity of the results of the two theories when applied to the tube-drawing operation is of special interest and convenience, because it suggests that the simpler analytical treatment might also be applied with some confidence to other cases of drawing, such as radial drawing from a.

flat bank, where the more accurate theory is intractable. It will be seen that the agreement between the theories extends at least to reductions of 60 per cent., which amply covers the useful range of radial drawing.

The fact that the simpler theory cannot be applied with equal confidence to the tube-sinking operation is not of serious consequence, because this operation has no useful analogue in pressing from a flat blank.

Fig. 14.



Drawing strains: + Mild steel. ⊕ Brass. ● Aluminium.

COMPARISON WITH EXPERIMENTAL RESULTS.

There are few experimental measurements with which these theoretical results can be properly compared, for there is little published work on the reduction of tubes and none apparently on pressure sinking. Stress

measurements are difficult to compare because correlation with theory requires a knowledge of the plastic stress-strain characteristics of the material, which have not been recorded. Measurements of thickness changes in the tube walls have been made, but, since the purpose of published experiments has usually (and quite properly) been to examine the effect of wall thickness, most measurements have been made on tubes whose walls are sufficiently thick to introduce a significant bending effect and are, therefore, unsuited for comparison with the present theory.

Of the available measurements on thin-walled tubes those recorded by Pomp and Schyller (1934), for a thickness ratio $t_0/d_0=0.03$, vary over a wide range and are scarcely comparable, since they refer to hot-drawn tubes. From Baldwin and Howald's (1944) results those obtained with thin-walled brass and copper tubes ($t_0/d_0<0.02$) indicate a wall thickening of about 4 per cent. for 10 per cent. and 50 per cent. reductions, and 8-10 per cent. for 25 per cent. reduction. These values agree quite well with the theoretical curve in fig. 9 corresponding to $c=m=1$.

Experimental work is in progress in the author's laboratory on the reduction of tubes $4\frac{1}{4}$ in. external diameter with walls $\frac{1}{8}$ in. thick. This investigation covers a wide range and is incomplete, but sufficient strain measurements have been made to afford comparison with theory, and these are plotted in fig. 14 together with theoretical curves for certain representative values of c and m . It will be seen that in all cases the measured strains diverge from the frictionless, non-hardening curve in the direction predicted by theory, and they show trends bearing at least a family resemblance to those of the theoretical curves in the figures.

The strain measurements plotted in fig. 14 are due to Messrs. G. C. Briggs and B. M. Botros, whose ready and careful assistance the author wishes to acknowledge.

REFERENCES.

- W. M. BALDWIN, and T. S. HOWARD, 1944, *Trans. A.S.M.*, **33**, 88.
POMP, and SCYLLER, 1934, *Mitt. K.W.I. Düss.*, **16**, 31.
G. SACHS, and W. M. BALDWIN, 1946, *Trans. A.S.M.E.*, **68**, 655.

LXXXII. *Edge Dislocations in Anisotropic Materials.*

J. D. ESHELBY,

H. H. Wills Physical Laboratory, University of Bristol*.

[Received May 18, 1949.]

ABSTRACT.

Burgers has discussed dislocations in materials with cubic symmetry for the general case in which the dislocation axis is an arbitrary curve. The results of the present paper are limited to dislocations of edge type with an infinite straight line as axis, but apply to a material with the symmetry of any of the crystal classes. The axis of the dislocation may be arbitrarily inclined to the symmetry axes of the material. Expressions are given for the elastic displacements and the energy of the dislocation. Nabarro's calculation of the width of a dislocation is extended to the anisotropic case.

§ 1. INTRODUCTION.

BURGERS (1939) has discussed elastic dislocations in both isotropic materials and crystals with cubic symmetry for the general case in which the axis of a dislocation is allowed to be an arbitrary curve. It would be difficult to extend his analysis to cover crystals of lower symmetry. However, for a dislocation of edge type whose axis is a straight line of infinite extent the problem is one of plane strain. Green (1945) has already solved the problem of generalized plane stress in an anisotropic plate. In the present paper Green's results are adapted to the case of plane strain (§ 2) and used to find the elastic displacements surrounding the dislocation (§ 3). In § 4 the elastic energy of the dislocation is calculated and in § 5 Nabarro's calculation of the width of a dislocation is repeated for the anisotropic case. § 6 contains some illustrative numerical results.

§ 2. THE SOLUTION OF PROBLEMS OF PLANE STRAIN IN ANISOTROPIC MATERIALS.

A problem in generalized plane stress in an isotropic plate can be solved by regarding it as a problem in plane strain and in the results replacing the stresses by their mean values across the plate and also replacing λ by $\lambda' = 2\lambda\mu/(\lambda + 2\mu)$, where λ and μ are the Lamé constants (Love 1944). Conversely a problem in plane strain could be replaced by one in plane stress. This latter process can be extended to the anisotropic case.

* Communicated by Prof. N. F. Mott, F.R.S.

For plane strain in which the z -displacement w vanishes and the x - and y -displacements u, v are independent of z we have, with the usual notation :

$$\begin{aligned} e_{zz} &= s_{31}\widehat{xx} + s_{32}\widehat{yy} + s_{33}\widehat{zz} + s_{34}\widehat{yz} + s_{35}\widehat{xz} + s_{36}\widehat{xy} = 0, \\ e_{yz} &= s'_{41}\widehat{xx} + s_{42}\widehat{yy} + s_{43}\widehat{zz} + s_{44}\widehat{yz} + s_{45}\widehat{xz} + s_{46}\widehat{xy} = 0, \\ e_{xz} &= s_{51}\widehat{xx} + s_{52}\widehat{yy} + s_{53}\widehat{zz} + s_{54}\widehat{yz} + s_{55}\widehat{xz} + s_{56}\widehat{xy} = 0. \end{aligned}$$

These equations can be used to eliminate $\widehat{zz}, \widehat{yz}, \widehat{xz}$ from the expressions for e_{xx}, e_{yy}, e_{xy} , which become

$$\begin{aligned} e_{xx} &= s'_{11}\widehat{xx} + s'_{12}\widehat{yy} + s'_{16}\widehat{xy}, \\ e_{yy} &= s'_{21}\widehat{xx} + s'_{22}\widehat{yy} + s'_{26}\widehat{xy}, \\ e_{xy} &= s'_{61}\widehat{xx} + s'_{62}\widehat{yy} + s'_{66}\widehat{xy}, \end{aligned}$$

where

$$s'_{ij} = \begin{vmatrix} s_{ij} & s_{i3} & s_{i4} & s_{i5} \\ s_{3j} & s_{33} & s_{34} & s_{35} \\ s_{4j} & s_{43} & s_{44} & s_{45} \\ s_{5j} & s_{53} & s_{54} & s_{55} \end{vmatrix} \div \begin{vmatrix} s_{33} & s_{34} & s_{35} \\ s_{43} & s_{44} & s_{45} \\ s_{53} & s_{54} & s_{55} \end{vmatrix}, \quad i, j = 1, 2, 6.$$

Clearly $s'_{ij} = s'_{ji}$.

The following form is convenient for calculation (Schwein's expansion, cf. Aitken (1946)) :

$$\begin{aligned} s'_{ij} &= s_{ij} - s_{3j}s_{i3}/s_{33} - \begin{vmatrix} s_{3j} & s_{33} \\ s_{4j} & s_{43} \end{vmatrix} \cdot \begin{vmatrix} s_{i3} & s_{i4} \\ s_{33} & s_{34} \end{vmatrix} / \begin{vmatrix} s_{33} & s_{34} \\ s_{43} & s_{44} \end{vmatrix} \\ &- \begin{vmatrix} s_{3j} & s_{33} & s_{34} \\ s_{4j} & s_{43} & s_{44} \\ s_{5j} & s_{53} & s_{54} \end{vmatrix} \cdot \begin{vmatrix} s_{i3} & s_{i4} & s_{i5} \\ s_{33} & s_{34} & s_{35} \\ s_{43} & s_{44} & s_{45} \end{vmatrix} / \begin{vmatrix} s_{33} & s_{34} & s_{35} \\ s_{43} & s_{44} & s_{45} \\ s_{53} & s_{54} & s_{55} \end{vmatrix} \end{aligned} \quad \dots \quad (1)$$

Comparison with the results given by Green (1945) for the corresponding problem in generalized plane stress shows that it is only necessary to replace his s_{ij} by s'_{ij} to obtain solutions of a problem in plane strain. With this change the relevant part of his analysis is given below.

Introduce the two complex variables

$$z_1 = x + i\lambda_1 y, \quad z_2 = x + i\lambda_2 y$$

where λ_1, λ_2 are complex constants given implicitly in terms of the elastic constants by the following relation with $n=1$ or 2 :

$$\lambda_n = \frac{1 - \gamma_n - i\delta_n}{1 + \gamma_n + i\delta_n}, \quad \gamma_n = \frac{\alpha_n - 1}{\alpha_n + 1 + 2(\alpha_n - \frac{1}{4}k_n^2)^{\frac{1}{2}}}, \quad \delta_n = \frac{-k_n}{\alpha_n + 1 + 2(\alpha_n - \frac{1}{4}k_n^2)^{\frac{1}{2}}} \quad \dots \quad (2)$$

where

$$\alpha_1 + \alpha_2 + k_1 k_2 = (2s'_{12} + s'_{66})/s'_{22}, \quad \alpha_1 \alpha_2 = s'_{11}/s'_{22}$$

$$k_1 + k_2 = -2s'_{26}/s'_{22}, \quad k_1 \alpha_2 + k_2 \alpha_1 = -2s'_{16}/s'_{22}.$$

$\alpha_n - \frac{1}{4}k_n^2$ and α_n are real and positive and k_n, γ_n, δ_n are real.

For convenience we replace Green's f, g by f_1, f_2 . In the expressions below, Σ represents summation over the values $n=1, 2$ and c.c. the complex conjugate of the quantity preceding it. To the stress function

$$\chi = \Sigma f_n(z_n) + \text{c.c.}$$

there correspond the displacements

$$u = -\Sigma C_n f'_n(z_n) + \text{c.c.}, \quad v = i\Sigma D_n f'_n(z_n) + \text{c.c.}$$

where

$$C_n = s'_{11}\lambda_n^2 - s'_{12} + i\lambda_n s'_{16},$$

$$D_n = \lambda_n^{-1}(s'_{12}\lambda_n^2 - s'_{22} + i\lambda_n s'_{26}),$$

and the stresses

$$\widehat{xx} = -\Sigma \lambda_n^2 f''_n(z_n) + \text{c.c.}$$

$$\widehat{yy} = \Sigma f''_n(z_n) + \text{c.c.}$$

$$\widehat{xy} = -i\Sigma f''_n(z_n) + \text{c.c.}$$

The couple on a circuit in the material is the change of

$$M = \Sigma \{z_n f'_n(z_n) - f_n(z_n)\} + \text{c.c.}$$

on going round the circuit. Similarly the x and y -components of the total force acting on the circuit are given by the changes of

$$X = -i\Sigma \lambda_n f'_n(z_n) + \text{c.c.}$$

$$Y = \Sigma f'_n(z_n) + \text{c.c.}$$

on going round it.

§3. EDGE DISLOCATIONS IN ANISOTROPIC BODIES.

To represent an edge dislocation we need, following Burgers (1939), a solution for which (u, v) increases by a constant vector \mathbf{b} on describing a circuit about the origin. The total force and couple on any circuit must vanish. Let $f_n(z_n) = \frac{1}{2}A_n \log z_n$, where A_n is a complex constant.

If the suffixes r, i denote the real and imaginary parts of a quantity then

$$\left. \begin{aligned} u &= -\Sigma \{ (C_{nr}A_{nr} - C_{ni}A_{ni}) \log \sqrt{(x_n^2 + y_n^2)} - (C_{ni}A_{nr} + C_{nr}A_{ni}) \tan^{-1}(y_n/x_n) \}, \\ v &= -\Sigma \{ (D_{ni}A_{nr} + D_{nr}A_{ni}) \log \sqrt{(x_n^2 + y_n^2)} + (D_{nr}A_{nr} - D_{ni}A_{ni}) \tan^{-1}(y_n/x_n) \}, \end{aligned} \right\} \quad \dots (3)$$

where

$$x_n = x - \lambda_{ni}y, \quad y_n = \lambda_{nr}y.$$

The couple given by M vanishes for any choice of A_1, A_2 . The condition that the force given by Y should vanish is $A_{2i} = -A_{1i}$, and the corresponding condition for X is

$$\lambda_{1r}A_{1r} + \lambda_{2r}A_{2r} - (\lambda_{1i} - \lambda_{2i})A_{1i} = 0. \quad \dots \dots \dots (4)$$

The change in (u, v) round a circuit about the origin arises from the change of 2π in the inverse tangents in (3). If the Burgers vector is required to be (b_x, b_y) we must have

$$C_{1i}A_{1r} + C_{2i}A_{2r} + (C_{1r} - C_{2r})A_{1i} = b_x/2\pi, \quad \dots \dots \dots (5)$$

$$D_{1r}A_{1r} + D_{2r}A_{2r} - (D_{1i} - D_{2i})A_{1i} = b_y/2\pi. \quad \dots \dots \dots (6)$$

Equations (4), (5) and (6) fix the values of the constants A_{1r}, A_{1i}, A_{2i} . In what follows we shall put $b_x = b, b_y = 0$, so that the Burgers vector lies along the x -axis: the slip-plane is then the plane $y = 0$. The expressions for (u, v) become fairly simple if $s'_{16} = s'_{26} = 0$. This covers a number of cases in which the slip-plane and Burgers vector are simply related to the symmetry axes of the material. Several cases have to be distinguished. If

$$s'_{16} = s'_{26} = 0 \text{ and } (2s'_{12} + s'_{66})^2 > 4s'_{11}s'_{22} \quad \dots \dots \dots (7)$$

λ_1 and λ_2 are real and

$$\left. \begin{aligned} u &= \frac{b}{2\pi} \frac{s'_{11}\lambda_1^2 - s'_{12}}{s'_{11}(\lambda_1^2 - \lambda_2^2)} \tan^{-1} \frac{\lambda_1 y}{x} + \frac{b}{2\pi} \frac{s'_{11}\lambda_2^2 - s'_{12}}{s'_{11}(\lambda_2^2 - \lambda_1^2)} \tan^{-1} \frac{\lambda_2 y}{x}, \\ v &= -\frac{b}{2\pi} \frac{s'_{12}\lambda_1 - s'_{22}\lambda_1^{-1}}{s'_{11}(\lambda_1^2 - \lambda_2^2)} \log \sqrt{(x^2 + y^2)} - \frac{b}{2\pi} \frac{s'_{22}\lambda_2 - s'_{22}\lambda_1^{-1}}{s'_{11}(\lambda_2^2 - \lambda_1^2)} \log \sqrt{(x^2 + y^2)}, \end{aligned} \right\} \quad \dots \dots \dots (8)$$

where

$$\lambda_1 = \lambda_{1r} = \alpha_1^{-\frac{1}{2}}, \quad \lambda_2 = \lambda_{2r} = \alpha_2^{-\frac{1}{2}}$$

and α_1, α_2 are the roots of

$$s'_{22}\alpha^2 - (2s'_{12} + s'_{66})\alpha + s'_{11} = 0.$$

On the other hand, if

$$s'_{16} = s'_{26} = 0 \text{ and } (2s'_{12} + s'_{66})^2 < 4s'_{11}s'_{22}, \quad \dots \dots \dots (9)$$

λ_1 and λ_2 are complex conjugates and

$$\left. \begin{aligned} u &= \frac{b}{4\pi} \left\{ \tan^{-1} \frac{\lambda_{1r}y}{x - \lambda_{1i}y} + \tan^{-1} \frac{\lambda_{1r}y}{x + \lambda_{1i}y} + \frac{s'_{11}(\lambda_{1r}^2 - \lambda_{1i}^2) - s'_{12}}{2s'_{11}\lambda_{1r}\lambda_{1i}} \right. \\ &\quad \times [-\log \sqrt{((x - \lambda_{1i}y)^2 + \lambda_{1r}^2y^2)} + \log \sqrt{((x + \lambda_{1i}y)^2 + \lambda_{1r}^2y^2)}] \Big\} \\ v &= -\frac{b}{4\pi} \left\{ \left[\frac{s'_{12}}{2s'_{11}\lambda_{1r}} + \frac{s'_{22}}{2s'_{11}\lambda_{1r}(\lambda_{1r}^2 + \lambda_{1i}^2)} \right] \left[\log \sqrt{((x - \lambda_{1i}y)^2 + \lambda_{1r}^2y^2)} \right. \right. \\ &\quad \left. \left. + \log \sqrt{((x + \lambda_{1i}y)^2 + \lambda_{1r}^2y^2)} \right] + \left[\frac{s'_{12}}{2s'_{11}\lambda_{1i}} - \frac{s'_{22}}{2s'_{11}\lambda_{1i}(\lambda_{1r}^2 + \lambda_{1i}^2)} \right] \right. \\ &\quad \left. \times \left[\tan^{-1} \frac{\lambda_{1r}y}{x - \lambda_{1i}y} - \tan^{-1} \frac{\lambda_{1r}y}{x + \lambda_{1i}y} \right] \right\}, \end{aligned} \right\} \quad (10)$$

where $\lambda_1 = \lambda_{1r} + i\lambda_{1i}$, can be found from (2) with

$$\alpha_1 = (s'_{11}/s'_{22})^{\frac{1}{2}}, \quad k_1^2 = 2(s'_{11}/s'_{22})^{\frac{1}{2}} - (2s'_{12} + s'_{66})/s'_{22}.$$

When $s'_{16} = s'_{26} = 0$ and $(2s'_{12} + s'_{66})^2 = 4s'_{11}s'_{22}$ (11)

$\lambda_{1r} = \lambda_{2r} = 1$ and $\lambda_{1i} = \lambda_{2i} = 0$. The displacements can be obtained by passing to the limit in (8) or (10). The result is

$$\left. \begin{aligned} u &= \frac{b}{2\pi} \tan^{-1} \frac{y}{x} + \frac{b}{2\pi} \frac{s'_{11} - s'_{12}}{2s'_{11}} \frac{xy}{x^2 + y^2}, \\ v &= -\frac{b}{2\pi} \frac{s'_{12} + s'_{22}}{2s'_{11}} \log \sqrt{(x^2 + y^2)} - \frac{b}{2\pi} \frac{s'_{12} - s'_{22}}{2s'_{11}} \frac{y^2}{x^2 + y^2}. \end{aligned} \right\} \quad (12)$$

For an isotropic substance

$$s'_{11}/s'_{22} = 1, \quad s'_{12}/s'_{11} = -\lambda/(\lambda + 2\mu)$$

and (12) agrees with the expression given by Burgers (1939).

In the general case when $s'_{16} \neq 0$, $s'_{26} \neq 0$ the expressions obtained by solving (4), (5) and (6) for A_{1r} , A_{2r} , A_{1i} and substituting them in (3) are rather clumsy. In the application in § 4, 5 only the values of (u, v) and \widehat{xy} at the slip-plane $y=0$ will be required. There is a discontinuity in u on crossing the x -axis. If

$$u_1(x) = \lim_{\epsilon \rightarrow 0} u(x, \epsilon), \quad u_2(x) = \lim_{\epsilon \rightarrow 0} u(x, -\epsilon)$$

and similarly for v_1, v_2 , then

$$\left. \begin{aligned} u_1 &= u_2 = P \log x, & x > 0, \\ u_1 &= P \log x + \frac{1}{2}b, & u_2 = P \log x - \frac{1}{2}b, & x < 0, \\ v_1 &= v_2 = Q \log x, & x > 0, \end{aligned} \right\} \quad (13)$$

where P, Q are functions of the s'_{ij} whose values will not be needed. If ϕ represents the discontinuity in u on crossing the x -axis then

$$\left. \begin{aligned} \phi &= u_1 - u_2 = 0, & x > 0, \\ &= b, & x < 0. \end{aligned} \right\} \quad (14)$$

The shear stress at the slip-plane is

$$\widehat{xy}_0 = \frac{b}{2\pi} \frac{K}{x} \quad (15)$$

where

$$K = \begin{vmatrix} \lambda_{1r} & \lambda_{2r} & -(\lambda_{1i} - \lambda_{2i}) \\ \lambda_{1i} & \lambda_{2i} & (\lambda_{1r} - \lambda_{2r}) \\ D_{1r} & D_{2r} & (D_{1i} - D_{2i}) \end{vmatrix} \div \begin{vmatrix} \lambda_{1r} & \lambda_{2r} & -(\lambda_{1i} - \lambda_{2i}) \\ C_{1i} & C_{2i} & (C_{1r} - C_{2r}) \\ D_{1r} & D_{2r} & (D_{1i} - D_{2i}) \end{vmatrix}.$$

When conditions (6) or (8) hold

$$K = [s'_{11}\{(2s'_{12} + s'_{66}) + 2(s'_{11}s'_{22})^{\frac{1}{2}}\}]^{-\frac{1}{2}} \quad (16)$$

This reduces to $K = 1/(2s'_{11})$ for the condition (10) and to $K = \mu/(1 - \sigma)$ for isotropy, where σ is Poisson's ratio.

§ 4. ELASTIC ENERGY OF THE DISLOCATION.

If the elastic energy density of a dislocation is integrated throughout the material the result diverges at the origin and infinity. The integration will therefore be confined to the space between a small cylinder of radius r_0 and a large cylinder of radius r_1 , both concentric with the z -axis. We can avoid having to know the stresses at all points of the medium by an artifice. Imagine an unstrained block of material to be cut along the half-plane $y=0, x \geq 0$. A dislocation may be formed by applying suitable forces to the cut surfaces and then welding them together again. The work done in this process is

$$W = \frac{1}{2} \int_{r_0}^{r_1} \{ \widehat{x} y_0 (u_1 - u_2) + \widehat{y} y_0 (v_1 - v_2) \} dx = \frac{b^2 K}{4\pi} \log \frac{r_1}{r_0} \quad (16)$$

per unit length of dislocation in the z -direction, by (14) and (15). This must be the energy of the dislocation per unit length. For isotropy it reduces to the value given by Koehler (1941, equation (15)).

§ 5. THE WIDTH OF THE DISLOCATION.

Near the origin the strains corresponding to the displacements (3) become large and Hooke's law cannot be expected to hold good. Nabarro (1940) has discussed the form of a dislocation near its centre assuming that isotropic elastic theory is valid except in the neighbourhood of the slip plane. The results of the present paper may be used to find the influence of anisotropy on his results.

Suppose that an infinite block of anisotropic material is cut in two along the plane $y=0$. Let arbitrary forces be applied to the surface of the upper half-block, uniformly distributed in the z -direction, so that the problem is a two-dimensional one. If exactly opposite forces are applied to corresponding points of the surface of the lower half-block the blocks can be welded together to give a dislocation of a more general kind than that already discussed: the discontinuity $\phi = u_1 - u_2$ varies from point to point along the x -axis.

Imagine a number of dislocations of the type (13) distributed along the x -axis. As we move along the axis in the negative direction ϕ changes by b each time a dislocation is passed. An arbitrary variation of ϕ with x can thus be considered as due to a continuous distribution of elementary dislocations along the x -axis. If the sum of the Burgers vectors of the elementary dislocations lying between x' and $x' + dx'$ is $b' dx'$ their contribution to the shear stress at $(x, 0)$ is, by (15)

$$\frac{b' K dx'}{2\pi(x-x')},$$

whilst

$$\frac{d\phi}{dx'} = -\frac{db'}{dx'};$$

By integration we find

$$\widehat{x} y_0 = \frac{K}{2\pi} \int_{-\infty}^{\infty} \frac{1}{x' - x} \frac{d\phi}{dx'} dx' \quad . \quad . \quad . \quad (17)$$

as the relation between the shear stress at any point and the relative displacements at all points of the slip-plane. Equation (17) corresponds to Nabarro's equation (10).

We now have to find a relation between ϕ and \widehat{xy}_0 at any point expressing the departure from Hooke's law. Suppose that the material is given a shear in which the atomic planes parallel to the slip-plane slide over each other without change of the spacing a between them. The only non-vanishing strain component is $e_{xy} = \partial u / \partial y$, so that if Hooke's law holds, $\widehat{xy} = s_{66}^{-1} \partial u / \partial y$. At the slip-plane this may be written

$$\widehat{xy} = \phi / s_{66} a. \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

For large ϕ/a this will cease to be true. Clearly \widehat{xy} must be periodic with period equal to the Burgers vector b which represents slip by one lattice spacing. If the relation between \widehat{xy} and ϕ is assumed to be sinusoidal and is required to reduce to (18) for small ϕ it must be

$$\widehat{xy} = \frac{1}{2\pi s_{66} a} \frac{b}{a} \sin \frac{2\pi\phi}{b}. \quad . \quad . \quad . \quad . \quad . \quad (19)$$

This corresponds to Nabarro's equation (1). An integral equation for ϕ can be found by equating the \widehat{xy}_0 of (17) to minus the \widehat{xy} of (19). (The normal to the surface of the upper half-block points in the negative direction.) The equation corresponding to Nabarro's (11) is thus

$$\int_{-\infty}^{\infty} \frac{1}{x-x'} \frac{d\phi}{dx'} dx' = \frac{b}{K a s_{66}} \sin \frac{2\pi\phi}{b}.$$

By comparison with Nabarro's equation (12),

$$\frac{\phi}{b} = -\frac{1}{\pi} \tan^{-1} \frac{x}{\zeta} \quad \text{with} \quad \zeta = \frac{1}{2} K s_{66} a. \quad . \quad . \quad . \quad (20)$$

ζ is a measure of the distance along the x -axis within which the departure from Hooke's law is important. It gives the "width" of the dislocation.

If $|x| \gg \zeta$, ϕ is $-\frac{1}{2}b$ when $x > 0$ and $\frac{1}{2}b$ when $x < 0$: for large x it thus differs from the ϕ of (14) by an inessential constant relative displacement of the upper and lower half blocks. The distribution of elementary dislocations which corresponds to (20) is

$$\frac{db'}{dx} = \frac{b}{\pi} \frac{\zeta}{x^2 + \zeta^2},$$

Hence from (13) we have

$$\begin{aligned} u_1, u_2 &= \pm \frac{1}{2} \phi + \frac{P b \zeta}{\pi} \int_{-\infty}^{\infty} \frac{\log |x-x'|}{x'^2 + \zeta^2} dx', \\ &= \mp \frac{b}{2\pi} \tan^{-1} \frac{x}{\zeta} + P b \log \sqrt{(x^2 + \zeta^2)}. \end{aligned}$$

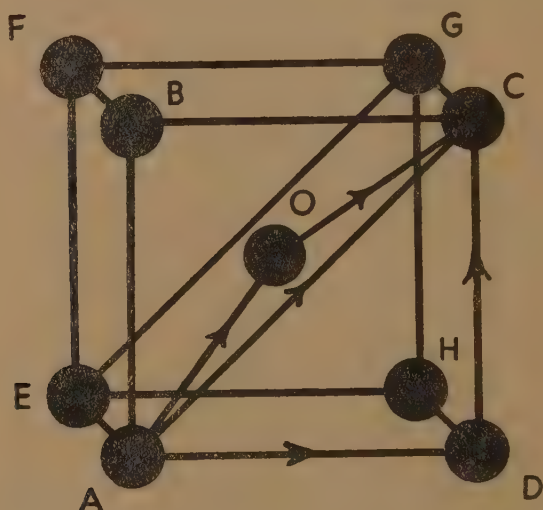
Similarly v_1 and v_2 are given by the second term of this expression with P replaced by Q .

The other results in Nabarro's paper which depend only on the displacement and shear stress at the slip-plane may similarly be found by comparison with his formulæ. For example, the ratio of the stress required to move a dislocation to the stress necessary to make the atomic planes move rigidly over one another is

$$4\pi(\zeta/b) \exp(-2\pi\zeta/b). \quad \dots \dots (21)$$

§ 6. DISCUSSION.

According to (16) the energy of a dislocation is proportional to the square of its Burgers vector and to the constant K . The logarithmic term will be neglected. If the body is treated as isotropic, K has the value $\mu/(1-\sigma)$ for all orientations and the energy depends only on the



Burgers vector. When anisotropy is taken into account K varies with the direction of the dislocation, since the s_{ij} change with the orientation of the coordinate axes.

The magnitude and direction of the Burgers vector and the direction of the slip-plane have to be introduced by considering the lattice structure of the crystal. The energy for various choices of Burgers vector and slip-plane can then be compared by using elastic theory. The quantity $\zeta/b = \frac{1}{2}Ks_{66}a/b$, which determines the ease of slipping, can also be calculated: K and s_{66} vary with orientation and a/b has values determined by the lattice geometry which may vary for different choices of Burgers vector and slip-plane.

As an example consider a crystal with body-centred cubic structure. The three shortest lattice displacements are AO , AD , AC . They are possible Burgers vectors. Three possible combinations of Burgers vectors and slip-plane are: (i) AD , $ADHE$; (ii) AC , $ACGE$; (iii) AO , $ACGE$.

It is convenient to write s_{ij}^0 for the values of the elastic constants when the x -, y - and z -axis are parallel to the fourfold axes of the material, and use s_{ij} to refer to an arbitrary coordinate system. The s_{ij}^0 are the elastic constants as usually tabulated.

For case (i)

$$K^{-2} = 2[(s_{11}^0)^2 - (s_{12}^0)^2] \{1 + (s_{12}^0/s_{11}^0) - 2(s_{12}^0/s_{11}^0)^2 + \frac{1}{2}(s_{44}^0/s_{11}^0)\} \quad (22)$$

$$s_{66} = s_{44}^0, \quad a/b = \frac{1}{2}AB/AD = \frac{1}{2}.$$

For case (ii) the s_{ij} have to be obtained from the s_{ij}^0 by the transformation corresponding to a rotation of 45° about AE. It will be found that K is still given by (22) whilst

$$s_{66} = 2(s_{11}^0 - s_{12}^0), \quad a/b = \frac{1}{2}BD/AC = \frac{1}{2}.$$

The non-vanishing s_{ij} for case (iii) are given in terms of the s_{ij}^0 by the relations

$$s_{11} = s_{11}^0 - \frac{2}{3}S, \quad s_{22} = s_{33} = s_{11}^0 - \frac{1}{2}S, \quad s_{44} = s_{44}^0 + \frac{2}{3}S, \quad s_{55} = s_{66} = s_{44}^0 + \frac{4}{3}S,$$

$$s_{12} = s_{13} = s_{12}^0 + \frac{1}{3}S, \quad s_{23} = s_{12}^0 + \frac{1}{6}S, \quad s_{25} = -s_{35} = \frac{1}{2}s_{46} = -\frac{1}{3}\sqrt{2}S,$$

$$\text{where } S = s_{11}^0 - s_{12}^0 - \frac{1}{2}s_{44}^0.$$

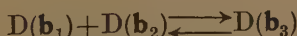
K can be found from (1) and (16). The value of a/b is $\frac{1}{2}BD/AD = \sqrt{(\frac{2}{3})}$.

Numerical values for α -iron are given in Table I., using the elastic constants given by Schmid and Boas (1935). The isotropic values of K and s_{66} are calculated from the elastic constants found by Schmid and Boas by averaging the anisotropic constants: this gives a fairer comparison than would the observed bulk values.

TABLE I.
Energy per unit length of a dislocation in α -iron.

	Slip-plane	Slip direction	$K \times 10^{-11}$ dyne/cm. ²	$s_{66} \times 10^{13}$ cm. ² /dyne	Energy $\times 10^5$ erg/cm.	a/b	ζ/b
(i)	(100)	[010]	10.5	8.6	6.82	$\frac{1}{2}$	0.23
(ii)	(110)	[010]	10.5	20.8	13.65	$\frac{1}{2}$	0.55
(iii)	(110)	[111]	12.2	16.7	5.95	$\sqrt{(\frac{2}{3})}$	0.83
(iv)	(Isotropy)		11.4	13.1	—	—	—

Frank (1949) has suggested that a dislocation may "dissociate" into two or more dislocations whose Burgers vectors added vectorially are equivalent to that of the original dislocation; conversely, several dislocations might "associate" to form a single one. If we use $D(\mathbf{b})$ to denote the dislocation with Burgers vector \mathbf{b} we may write the symbolical reaction



where the \mathbf{b} s satisfy

$$\mathbf{b}_1 + \mathbf{b}_2 = \mathbf{b}_3$$

and the axes of the dislocations are perpendicular to the plane in which $\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3$ lie. According to (16) the energy of dissociation is

$$E = (K_1 b_1^2 + K_2 b_2^2 - K_3 b_3^2) / 4\pi$$

per unit length. The three K s have to be calculated for the orientation of the corresponding \mathbf{b} , as explained above. The reaction will go in the direction of dissociation or association according as E is positive or negative. If the medium is isotropic the condition becomes

$$b_1^2 + b_2^2 \gtrless b_3^2,$$

i.e. there will be dissociation or association according as the angle between \mathbf{b}_1 and \mathbf{b}_2 is acute or obtuse.

For a body-centred material the reaction



has $E=0$ in the case of isotropy, since ADC is a right angle. The same is true when anisotropy is taken into account, since according to the previous discussion all three K s are given by (22). The reaction



has a negative E in the isotropic case, since AOC is obtuse. Whether E is positive or negative in the anisotropic case will depend on the elastic constants of the particular material. For example, in the case of α -iron taking $K=11.4 \times 10^{11}$ (from the entry (iv) of Table I.) and $AD=2.86 \text{ \AA}$ gives E (isotropic) $= -3.7 \times 10^5 \text{ erg/cm.}$, whilst entries (ii) and (iii) give directly E (anisotropic) $= (2 \times 5.95 - 13.65) \times 10^{-5} = -1.75 \times 10^{-5} \text{ erg/cm.}$ In this case taking into account anisotropy reduces the magnitude of E , but is insufficient to reverse its sign.

The last column of Table I. implies, in conjunction with equation (21), that of the three cases considered dislocations of type (iii) (with Burgers vectors along cube diagonals) should move much more readily under the influence of an applied shear stress than types (i) and (ii).

Similar calculations could be made for other materials and other crystal structures. The necessary calculation of the s_{ij} from the s_{ij}^0 and the s'_{ij} from the s_{ij} are, however, rather tedious.

REFERENCES.

- AITKEN, A. C., 1946, *Determinants and Matrices* (Edinburgh: Oliver and Boyd), p. 109.
 BURGERS, J. M., 1939, *Proc. Kon. Ned. Akad. v. Wet.*, **42**, 293, 378.
 FRANK, F. C., 1949, Private communication.
 GREEN, A. E., 1945, *Proc. Camb. Phil. Soc.*, **41**, 224.
 KOEHLER, J. S., 1941, *Phys. Rev.*, **60**, 397.
 LOVE, A. E. H., 1944, *Elasticity* (Cambridge: University Press).
 NABARRO, F. R. N., 1940, *Proc. Phys. Soc.*, **52**, 34.
 SCHMID, E., and BOAS, W., 1935, *Kristallplastizität* (Berlin: Springer), pp. 200, 326.

where R denotes the real part of the complex function and f_1, f_2 are arbitrary functions. $k = \sqrt{(E_1/E_2)}$, E_1, E_2 denote the Young's moduli in x, y directions respectively.

Equation (3) can be written as by virtue of (1)

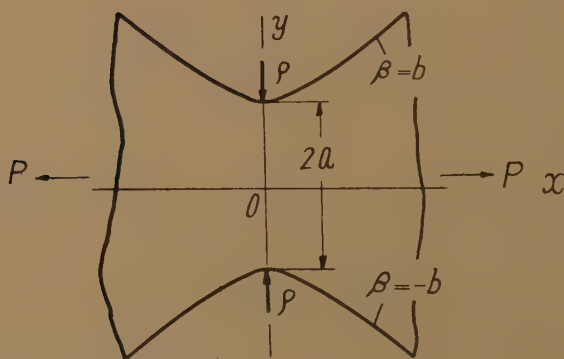
$$\chi = R[\phi_1(\alpha + i\beta) + \phi_2(\alpha' + i\beta')], \quad (4)$$

where ϕ_1, ϕ_2 are arbitrary functions.

The stress components of this case are

$$\left. \begin{aligned} \widehat{\alpha\alpha} &= \frac{2}{c^2(\cosh 2\alpha + \cos 2\beta)} \left\{ \frac{\partial^2 \chi}{\partial \beta^2} + \frac{\sin 2\beta}{(\cosh 2\alpha + \cos 2\beta)} \frac{\partial \chi}{\partial \beta} \right. \\ &\quad \left. + \frac{\sinh 2\alpha}{(\cosh 2\alpha + \cos 2\beta)} \frac{\partial \chi}{\partial \alpha} \right\}, \\ \widehat{\beta\beta} &= \frac{2}{c^2(\cosh 2\alpha + \cos 2\beta)} \left\{ \frac{\partial^2 \chi}{\partial \alpha^2} - \frac{\sinh 2\alpha}{(\cosh 2\alpha + \cos 2\beta)} \frac{\partial \chi}{\partial \alpha} \right. \\ &\quad \left. - \frac{\sin 2\beta}{(\cosh 2\alpha + \cos 2\beta)} \frac{\partial \chi}{\partial \beta} \right\}, \\ \widehat{\alpha\beta} &= -\frac{2}{c^2(\cosh 2\alpha + \cos 2\beta)} \left\{ \frac{\partial^2 \chi}{\partial \alpha \partial \beta} - \frac{\sinh 2\alpha}{(\cosh 2\alpha + \cos 2\beta)} \frac{\partial \chi}{\partial \beta} \right. \\ &\quad \left. + \frac{\sin 2\beta}{(\cosh 2\alpha + \cos 2\beta)} \frac{\partial \chi}{\partial \alpha} \right\}. \end{aligned} \right\} (5)$$

Fig. 1.



We shall treat the problem of a notched plate stretched by opposite forces in x direction (see Fig. 1). This problem for the case of an isotropic one was treated formerly by H. Neuber (1933).

Now, putting

$$\left. \begin{aligned} \phi_1(\alpha + i\beta) &= Ac \cosh(\alpha + i\beta) + Bc(\alpha + i\beta) \sinh(\alpha + i\beta) \\ &\quad - iCc \sinh(\alpha + i\beta), \\ \phi_2(\alpha' + i\beta') &= A'c' \cosh(\alpha' + i\beta') - Bc'(\alpha' + i\beta') \sinh(\alpha' + i\beta'), \end{aligned} \right\}$$

and inserting the expressions of ϕ_1 and ϕ_2 into equation (4), the stress function becomes

$$\begin{aligned}\chi = & Ac \cosh \alpha \cos \beta + A'c' \cosh \alpha' \cos \beta' \\ & + Bc(\alpha \sinh \alpha \cos \beta - \beta \cosh \alpha \sin \beta) \\ & - Bc'(\alpha' \sinh \alpha' \cos \beta' - \beta' \cosh \alpha' \sin \beta') \\ & + Cc \cosh \alpha \sin \beta. \quad \dots \dots \dots (6)^*\end{aligned}$$

Since the periphery is free from traction, $\widehat{\beta\beta}$ and $\widehat{\alpha\beta}$ must vanish at $\beta = \pm b$. If we determine χ to satisfy the following conditions:

$$\chi = \text{const.}, \quad \frac{\partial \chi}{\partial \beta} = 0, \quad \dots \dots \dots (7)$$

then, at the boundary the above conditions are fully satisfied. The condition that χ vanishes at the boundary is

$$A + A' - B(b - kb') \tan b + C \tan b = 0,$$

and the condition that $\frac{\partial \chi}{\partial \beta}$ vanishes at the boundary is

$$\begin{aligned}A \tan b + A'k^2 \tan b + B\{b - b'k + (1 - k^2) \tan b\} - C &= 0, \\ Ak^2 \tan^2 b + (1 - k^2 + k^2 \tan^2 b)A' + B(k^2b \tan b - k^3b' \tan b + k^2 - 1) \\ - Ck^2 \tan b &= 0.\end{aligned}$$

To satisfy the above conditions, the following relations must hold between the arbitrary constants:

$$A = -A' = -B. \quad \dots \dots \dots (8)$$

From eq. (5), the normal stress on the y axis is

$$\widehat{\alpha\alpha} = A \left(\frac{1}{c \cos \beta} - \frac{k^2}{c' \cos \beta'} \right).$$

If we denote the resultant force acting in x direction by P , then

$$2 \int_0^a \widehat{\alpha\alpha} dy = P, \quad \text{or} \quad A = \frac{P}{2(b - kb')}. \quad \dots \dots \dots (9)$$

Now the arbitrary constants are determined and the stress function and therefore the stress components are all obtained.

The normal stress on the y axis is given by

$$\widehat{\alpha\alpha} = \frac{P}{2(b - kb')} \left(\frac{1}{c \cos \beta} - \frac{k^2}{c' \cos \beta'} \right). \quad \dots \dots \dots (10)$$

and the maximum stress is

$$[\widehat{\alpha\alpha}]_{\max} = \frac{P(1 - k^2)}{2c(b - kb') \cos b}. \quad \dots \dots \dots (11)$$

* The last term of the right side of eq. (6) is independent of the stress distribution.

If we denote by α_c the stress concentration factor, the ratio of the maximum stress to the mean normal stress on the y axis, we have

$$\alpha_c = \frac{(1-k^2) \tan b}{b-kb'} \quad \dots \quad (12)$$

If we use the expression of χ shown in footnote ‡, then by a similar calculation we can readily obtain the stress concentration factor as

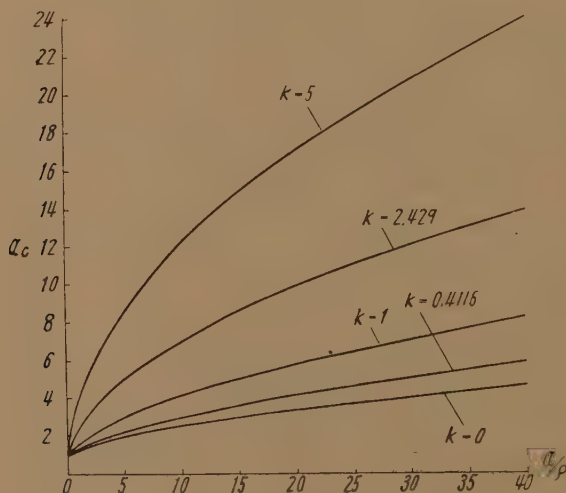
$$\alpha'_c = \frac{(k_1^2 - k_2^2) \tan b}{k_1 b' - k_2 b''}, \quad \dots \quad (13)$$

where

$$\tan b' = k_1 \tan b, \quad \tan b'' = k_2 \tan b.$$

In the case of an oak plate, placed the direction of the fibres of wood parallel to the axis of x , $k_1 = 2.31$, $k_2 = 1.05$, $k = 2.429$.

Fig. 2.



When the ratio of a to the radius of curvature ρ at the bottom of the notch is 4, eqs. (12) and (13) give the stress concentration factors as

$$\alpha_c = 4.42, \quad \alpha'_c = 4.34.$$

We can see from the result that the discrepancy between the values given by both eqs. is comparatively small.

When k approaches to unity in eq. (12), we can find the stress concentration factor for an isotropic one

$$\alpha_c = \frac{2 \tan b}{b + \sin b \cos b}, \quad \dots \quad (14)$$

which coincides with the result obtained by H. Neuber. Fig. 2 shows the relation between the stress concentration factor and the radius of curvature at the bottom of the notch for several values of k . $k = 2.429$ or

0.4116 is the case of an oak plate cut parallel to the grain of wood. Experimental results show that in almost all wooden plates, k is smaller than 5, so from the curves in fig. 2 we can readily estimate the approximate value of k of all kinds of wooden plates.

REFERENCES.

- NEUBER, H., 1933, *Z. angew. Math. Mech.*, **13**, 439.
 OKUBO, H., 1936, *Science Reports of the Tōhoku Imperial University* (March);
 1939, *Phil. Mag.* [7], **27**, 508; 1941, *Z. angew. Math. Mech.*, **21**, 000;
 1942, *J. Soc. Mech. Eng., Japan*, **45**, 589; 1944, *Trans. Soc. Mech. Eng., Japan*, **10**, I-176; 1945, *Ibid.*, **11**, I-23.

LXXXIV. Distance Correlations in an Ideal Fermi-Dirac Gas.

By P. L. BHATNAGAR and K. S. SINGWI.

University of Delhi, Delhi, India *.

[Received April 21, 1949.]

§ 1.

THE peculiar distance correlations between the molecules of an ideal Fermi-Dirac or Bose-Einstein gas are due to the symmetry properties of the wave-functions. Uhlenbeck and Gropper (1932) have shown that in the limiting case of high temperatures and low densities, *i. e.* $V \gg N\lambda^3$, the probability of finding a particle within the distance interval $(r, r+dr)$ of a chosen particle is

$$\frac{N}{V} \{1 \pm \exp(-2\pi r^2/\lambda^2)\} 4\pi r^2 dr, \quad \dots \dots \dots (1)$$

where λ is the de Broglie wavelength corresponding to the temperature T , given by

$$\lambda = \frac{h}{(2\pi m k T)^{\frac{1}{2}}}, \quad \dots \dots \dots (2)$$

and where the $+$ sign characterizes the Bose-Einstein case and the $-$ sign characterizes the Fermi-Dirac case. N denotes the number of particles in the volume V . For strong degeneracy, *i. e.* $V \leq N\lambda^3$, in the Fermi-Dirac case the corresponding probability expression as obtained by Wigner and Seltz (1933) is

$$\frac{N}{V} \left\{ 1 - 9 \left(\frac{2\pi k_0 r \cos 2\pi k_0 r - \sin 2\pi k_0 r}{(2\pi k_0 r)^3} \right)^2 \right\} 4\pi r^2 dr, \quad \dots \dots (3)$$

where k_0 , the wave number corresponding to the top of the Fermi distribution is given by

$$\frac{4\pi}{3} k_0^3 = \frac{N}{V}. \quad \dots \dots \dots (4)$$

* Communicated by the Authors.

In this note we will explicitly calculate this correlation effect for a Fermi-Dirac gas at all temperatures, and will derive formulæ (1) and (3) to a higher degree of approximation. The procedure is elegant.

§2.

Following London (1943), it can easily be shown that, in the Fermi-Dirac case, the mean number of particles per unit volume, or the particle density, at a distance r of any given particle is given by

$$D(r) = \frac{N}{V} \left[1 - \frac{1}{N^2} \left\{ \left(\sum_e \bar{n}_e \frac{\sin 2\pi k_e r}{2\pi k_e r} \right)^2 - \sum_e \bar{n}_e^2 \right\} \right], \quad \dots \quad (5)$$

where

$$\bar{n}_e = \frac{1}{\zeta^{-1} \exp(\epsilon_e/\kappa T) + 1}, \quad \dots \quad (6)$$

and

$$\zeta = \exp(\eta/\kappa T), \quad \dots \quad (7)$$

η being the Gibb's free energy per particle.

The non-degenerate case is characterized by $\zeta \leq 1$ and the degenerate case by $\zeta \gg 1$. We shall consider these two cases separately.

Case 1 :—Non-degenerate $\zeta \leq 1$.

$$\begin{aligned} \sum_e \bar{n}_e \frac{\sin 2\pi k_e r}{2\pi k_e r} &= \int_0^\infty \frac{\sin 2\pi k r}{2\pi k r} \frac{4\pi g V k^2 dk}{\zeta^{-1} \exp(k^2 \hbar^2/2m\kappa T) + 1} \\ &\equiv I_1, \end{aligned}$$

where g is the weight factor of the particle, $g=2$ for electron. Putting

$$x = \frac{k\hbar}{\sqrt{(2m\kappa T)}},$$

and using (2), we have

$$\begin{aligned} I_1 &= \int_0^\infty \frac{2gV}{\pi\lambda^2 r} \cdot \frac{\sin \frac{2\pi^{\frac{1}{2}} r x}{\lambda}}{\zeta^{-1} \exp x^2 + 1} x dx \\ &= \frac{2gV}{\pi r \lambda^2} \sum_{n=1}^\infty (-1)^{n+1} \zeta^n \int_0^\infty x \sin \left(\frac{2\pi^{\frac{1}{2}} r x}{\lambda} \right) \exp(-nx^2) dx \\ &= \frac{gV}{\lambda^3} \sum_{n=1}^\infty \frac{(-1)^{n+1}}{n^{3/2}} \zeta^n \exp(-\pi r^2/n\lambda^2). \quad \dots \quad (8) \end{aligned}$$

Also

$$\begin{aligned} \sum_e \bar{n}_e^2 &= \int_0^\infty \frac{4\pi g V k^2 dk}{\{\zeta^{-1} \exp(k^2 \hbar^2/2m\kappa T) + 1\}^2} \\ &= \frac{4gV}{\pi^{\frac{1}{2}} \lambda^3} \int_0^\infty \frac{x^2 dx}{\{\zeta^{-1} \exp x^2 + 1\}^2} \\ &= \frac{gV}{\lambda^3} \sum_{n=1}^\infty (-1)^{n+1} \frac{n}{(n+1)^{3/2}} \zeta^{n+1}. \quad \dots \quad (9) \end{aligned}$$

Substituting (8) and (9) in (5), we have

$$\begin{aligned}
 D(r) &= \frac{N}{V} \left[1 - \frac{1}{N^2} \left\{ \left(\frac{gV}{\lambda^2} \sum_{n=1}^{\infty} \frac{(-1)^{n+1} \zeta^n}{n^{3/2}} \exp \left(-\frac{\pi r^2}{n\lambda^2} \right) \right)^2 \right. \right. \\
 &\quad \left. \left. - \frac{gV}{\lambda^3} \sum_{n=1}^{\infty} \frac{(-1)^{n+1} n}{(n+1)^{3/2}} \zeta^{n+1} \right\} \right] \\
 &= \frac{N}{V} \left[1 - \exp \left(\frac{-2\pi r^2}{\lambda^2} \right) + \frac{1}{2^{\frac{1}{2}}} \zeta \exp \left(\frac{-3\pi r^2}{2\lambda^2} \right) - \zeta^2 \left\{ \frac{1}{2^3} \exp \left(-\frac{\pi r^2}{\lambda^2} \right) \right. \right. \\
 &\quad \left. \left. + \frac{1}{2^{\frac{1}{2}}} \exp \left(\frac{-4\pi r^2}{3\lambda^2} \right) \right\} + \frac{1}{N} \left(\frac{1}{2^{3/2}} \zeta - \frac{2}{3^{3/2}} \zeta^2 \right) + O(\zeta^3) \right], \quad \dots \quad (10)
 \end{aligned}$$

where use has been made of the relation

$$\zeta = \frac{N\lambda^3}{gV}.$$

The last but one term in (10) is negligible since N is large. The dependence of η on temperature is given by (Kothari and Singh 1941)

$$\eta = \kappa T [\log A_0 + 2b_2 A_0 - \frac{3}{2} b_3 A_0^2 + \dots], \quad \dots \quad (11)$$

where

$$A_0 = \frac{4}{3\pi^{\frac{1}{2}}} \left(\frac{\eta_0}{\kappa T} \right)^{3/2},$$

η_0 being the top of the Fermi distribution at absolute zero,

$$b_2 = \frac{1}{2^{5/2}}, \quad b_3 = \frac{2}{3^{5/2}} - \frac{1}{4^{3/2}}.$$

Hence

$$\begin{aligned}
 \zeta &= \exp(\eta/\kappa T) \\
 &= \frac{4}{3\pi^{\frac{1}{2}}} (\eta_0/\kappa T)^{3/2} + \frac{2^{5/2}}{9\pi} (\eta_0/\kappa T)^3,
 \end{aligned}$$

and

$$\zeta^2 = \frac{16}{9\pi} (\eta_0/\kappa T)^3,$$

neglecting powers of $\eta_0/\kappa T$ higher than the third. Substituting the above values of ζ and ζ^2 in (10) we have finally

$$\begin{aligned}
 D(r) &= \frac{N}{V} \left[1 - \exp \left(\frac{-2\pi r^2}{\lambda^2} \right) + \frac{1}{3} \left(\frac{8}{\pi} \right)^{\frac{1}{2}} (\eta_0/\kappa T)^{3/2} \exp \left(\frac{-3\pi r^2}{2\lambda^2} \right) \right. \\
 &\quad \left. + \frac{4}{9\pi} (\eta_0/\kappa T)^3 \left\{ \exp \left(\frac{-3\pi r^2}{2\lambda^2} \right) - \frac{1}{2} \exp \left(\frac{-\pi r^2}{\lambda^2} \right) \right. \right. \\
 &\quad \left. \left. - 2^{\frac{1}{2}} \exp \left(\frac{-4\pi r^2}{3\lambda^2} \right) \right\} \right]. \quad \dots \quad (12)
 \end{aligned}$$

To the zeroth order of approximation

$$D(r) = \frac{N}{V} \left[1 - \exp \left(\frac{-2\pi r^2}{\lambda^2} \right) \right],$$

which is same as (1).

Case II :—Degenerate $\zeta \gg 1$.

Substituting as before

$$x = \frac{kh}{(2m\kappa T)^{\frac{1}{2}}}$$

and using (2) and (6), we have

$$\begin{aligned} \Sigma_e \bar{n}_e \frac{\sin 2\pi k_e r}{2\pi k_e r} &= \frac{2gV}{\pi r \lambda^2} \int_0^\infty \frac{x \sin \frac{2\pi^{\frac{1}{2}} r x}{\lambda} dx}{\zeta^{-1} \exp(x^2) + 1} \\ &= \frac{gV}{\pi r \lambda^2} \int_0^\infty \frac{1}{\zeta^{-1} \exp(u) + 1} \frac{d\phi(u)}{du} du \quad \dots (13) \\ &\equiv I_2, \end{aligned}$$

where

$$\begin{aligned} u &= x^2, \quad \phi(u) = \int \sin(au^{\frac{1}{2}}) du, \\ &= \frac{2}{a^2} [\sin(au^{\frac{1}{2}}) - au^{\frac{1}{2}} \cos(au^{\frac{1}{2}})], \\ a &= \frac{2\pi^{\frac{1}{2}} r}{\lambda}, \end{aligned}$$

Since $\phi(0)=0$ and $\phi(u)$ is regular, we have, applying the Sommerfeld lemma to (13),

$$\begin{aligned} I_2 &= \frac{gV}{\pi r \lambda^2} \left[\frac{2}{a^2} \{ \sin(au_0^{\frac{1}{2}}) - (au_0^{\frac{1}{2}}) \cos(au_0^{\frac{1}{2}}) \} \right. \\ &\quad + 2C_2 \frac{a}{2} u_0^{-\frac{1}{2}} \cos(au_0^{\frac{1}{2}}) + 2C_4 \frac{a}{4} \left\{ \frac{3}{2} u_0^{-5/2} \cos(au_0^{\frac{1}{2}}) \right. \\ &\quad \left. \left. + \frac{3}{2} au_0^{-\frac{1}{2}} \sin(au_0^{\frac{1}{2}}) - \frac{a^2}{2} u_0^{-3/2} \cos(au_0^{\frac{1}{2}}) \right\} \right], \quad \dots (14) \end{aligned}$$

where

$$u_0 = \log \zeta = \eta / \kappa T, \quad C_2 = \frac{\pi^2}{12}, \quad C_4 = \frac{7\pi^4}{720}.$$

In the degenerate case the temperature dependence of η is given by (Kothari and Singh 1941)

$$\eta = \eta_0 \left[1 - C_2 \left(\frac{\kappa T}{\eta_0} \right)^2 - \frac{3}{4} (C_2^2 + C_4) \left(\frac{\kappa T}{\eta_0} \right)^4 + \dots \right]. \quad \dots (15)$$

Using (15) in (14) and retaining terms up to the second power of $\frac{\kappa T}{\eta_0}$, we have

$$\begin{aligned} I_2 &= \frac{Vg}{2\pi^2 r^3} \left[\{ \sin(b\eta_0^{\frac{1}{2}}) - (b\eta_0^{\frac{1}{2}}) \cos(b\eta_0^{\frac{1}{2}}) \} \right. \\ &\quad \left. \times C_2 \frac{b^2}{2} \eta_0 \left(\frac{\kappa T}{\eta_0} \right)^2 \{ (b\eta_0^{\frac{1}{2}}) \cos(b\eta_0^{\frac{1}{2}}) - \sin(b\eta_0^{\frac{1}{2}}) \} \right]. \quad \dots (16) \end{aligned}$$

Also

$$\begin{aligned}\Sigma \bar{n}_e^2 &= \frac{4gV}{\pi^{\frac{1}{2}}\lambda^3} \int_0^\infty \frac{x^2 dx}{[\zeta^{-1} \exp(x^2) + 1]^2} \\ &= \frac{2gV}{\pi^{\frac{1}{2}}\lambda^3} \int_0^\infty \frac{1}{\zeta^{-1} \exp(u) + 1} \frac{d\phi(u)}{du} du, \quad \dots \quad (17)\end{aligned}$$

where

$$\phi(u) = \int_0^u \frac{u^{\frac{1}{2}} du}{\zeta^{-1} \exp(u) + 1}.$$

Since $\phi(0)=0$ and $\phi(u)$ is regular, we have, applying the Sommerfeld lemma to (17),

$$\Sigma \bar{n}_e^2 = \frac{2gV}{\pi^{\frac{1}{2}}\lambda^3} \left[\int_0^{u_0} \frac{u^{\frac{1}{2}} du}{\zeta^{-1} \exp(u) + 1} + 2 \frac{C_2}{4} \left(\frac{1}{u_0^{\frac{1}{2}}} - u_0^{\frac{1}{2}} \right) + \dots \right]. \quad \dots \quad (18)$$

It is easily seen that

$$\int_0^{u_0} \frac{u^{\frac{1}{2}} du}{\zeta^{-1} \exp(u) + 1} < u_0^{3/2}$$

and

$$\frac{1}{N^2} \Sigma \bar{n}_e^2 = 0 \left(\frac{1}{N} \right),$$

which is, therefore, neglected in (5).

Substituting (16) in (5), and using the relations

$$N = \frac{4\pi gV}{3} k_0^3$$

and

$$\eta_0 = \frac{h^2}{2m} \left(\frac{3N}{4\pi gV} \right)^{2/3}$$

we have

$$D(r) = \frac{N}{V} \left[1 - 9 \left\{ \frac{\{1 - \frac{1}{2} C_2 (2\pi r k_0)^2 (\kappa T / \eta_0)^2\} \{\sin 2\pi r k_0 - 2\pi r k_0 \cos 2\pi r k_0\}^2}{(2\pi r k_0)^3} \right\}^2 \right] \quad \dots \quad (19)$$

which, when the temperature term is neglected, reduces to the Wigner-Seitz formula (3).

REFERENCES.

- KOTHARI and SINGH, 1941, *Proc. Roy. Soc. A*, **180**, 414.
 LONDON, F., 1943, *Journ. Chem. Phys.*, **11**, 203.
 UHLENBECK and GROPPER, 1932, *Phys. Rev.*, **41**, 79.
 WIGNER and SEITZ, 1933, *Phys. Rev.*, **43**, 804.

LXXXV. *The Calculation of the Optimum Parameters for a Following System.*

By C. MACK, M.A.*

[Received May 9, 1949.]

ABSTRACT.

This paper gives what is, in the author's opinion, the quickest way of calculating the optimum values of the parameters of a linear "following" system (such as most servomechanisms, video-amplifying circuits and any system where an output is designed to follow an input as quickly and as accurately as possible).

The method given here consists of calculating C , the integral of the square of the difference between input and output for different sets of parametric values (a comparatively simple formula can be found for C). The set which gives the minimum value of C is then chosen and the parameters varied slightly from their values at this point and other tests applied (of which the best is the E-test). The minimum value of C does not usually give the best following; the minimum value of E often gives a very satisfactory performance, but E is more troublesome to evaluate. Finally, in a few cases, the output has to be plotted graphically but as this involves solving algebraic equations it takes time, despite recent improved methods (Porter and Mack 1949); and to treat all cases by the plotting method would be extremely laborious (though with skill and experience it is possible to do much from the roots of the equations alone, *i. e.* by rejecting cases where the time-constants are long).

Other quantities which are useful such as (i) the frequency response and (ii) a stability test, can be worked out with little extra labour during the process.

§1. INTRODUCTION AND NOTATION.

We shall consider cases in which the output v is related to the input v_r by the Operational Calculus formula

$$v = \frac{a_0 + a_1 p + a_2 p^2 \dots + a_m p^m}{b_0 + b_1 p + b_2 p^2 \dots + b_n p^n} \cdot v_1 \equiv \frac{A(p)v_1}{B(p)}, \text{ say, } \dots \quad (1.1)$$

where $m \leq n$, $p \equiv d/dt$, $p^2 \equiv d^2/dt^2$, etc. and the a 's and b 's are functions of the parameters of the system.

For convenience we shall assume that $a_0 = b_0$. If this is not so, it means that the output does not follow the input but a multiple (a_0/b_0) of the input. We shall consider $(a_0 v_1 / b_0) = V$, say, to be the input in all further

* Communicated by Dr. F. C. Toy.

work and shall assume that $a_0=b_0$. This is equivalent to considering a fictitious system whose operation, however, bears a simple relation to the real system. With $a_0=b_0$ we further write

$$v = \frac{A(p)V}{B(p)} \quad \dots \dots \dots (1.2)$$

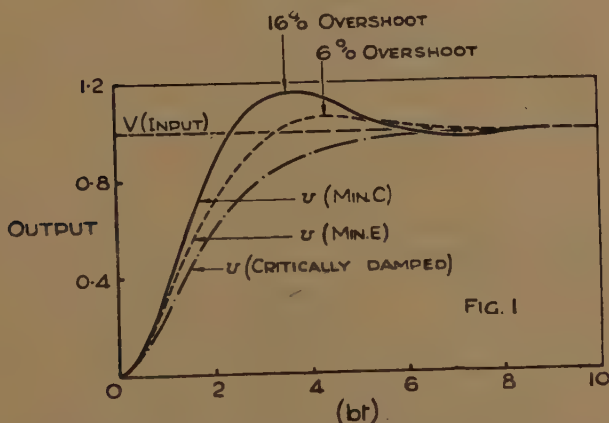
$$\frac{B(p)-A(p)}{B(p)} = p \frac{F(p)}{B(p)} = p \frac{[f_0+f_1p+f_2p^2+\dots]}{[b_0+b_1p+b_2p^2+\dots]} \quad \dots \quad (1.3)$$

where $f_r=b_{r+1}-a_{r+1}$, $r=0, 1, \dots (n-1)$. If $m < n$ some of the a 's will be zero.

In general v will not follow V exactly but will describe some curve similar to the curves of fig. 1 where V is a Heaviside Unit Function. Various quantities may be computed as a measure of the "goodness" of following. We shall be concerned with the following quantities:

$$C \equiv \int_0^\infty [V-v]^2 dt, \quad D \equiv \int_0^\infty [V-v] dt \quad \text{and} \quad E \equiv \int_0^\infty [t(V-v)]^2 dt.$$

Fig. 1.



It must be stated that none of the quantities are ideal and that the ultimate test is obtained by graphical plotting of the output v .

We consider a number of types of input in this paper, as different treatment is required in different cases.

§2. THE C-TEST FOR HEAVISIDE UNIT FUNCTION INPUT I.

Here we take $V=I$ (*i. e.* $V=0$ for $-\infty < t < 0$ and $V=1$ for $0 \leq t < \infty$). This is the input likely to be of most general interest. It is shown in the Appendix 8.1 that, provided $B(p)$ has no roots with positive real parts, when $V=I$,

$$\pi C = \int_0^\infty \frac{[f_0-f_2w^2+f_4w^4-\dots]^2 + w^2[f_1-f_3w^2+\dots]^2}{[b_0-b_2w^2+b_4w^3-\dots]^2 + w^2[b_1-b_3w^2+\dots]^2} dw \quad \dots (2.1)$$

$$= \int_0^\infty \frac{[f(w^2)]^2 + w^2[g(w^2)]^2}{[b(w^2)]^2 + w^2[c(w^2)]^2} dw, \text{ say. } \dots \dots \dots (2.2)$$

Though C is apparently an infinite integral it is fairly easily calculated. All the terms are positive ; for sufficiently large values of w the integrand approaches $f_{n-1}^2/(b_n w)^2$ and the integral from here to infinity is $f_{n-1}^2/(b_n^2 w)$. There is no need for high accuracy in the evaluation of $f(w^2)$, $g(w^2)$, $b(w^2)$ and $c(w^2)$ except when they all approach zero together (if $b(w^2)$ and $c(w^2) \rightarrow 0$ then the integrand will be large and C will be large too, an undesirable case) ; and they can be evaluated quickly and easily for simple values of w .

C is calculated for a number of sets of parametric values, and since the object is to make C a minimum *, a case in which C is obviously larger than some previous value can be left after a few values of the integrand have been found.

Finally, the trapezoidal rule of finding an area is good enough except when C is near its minimum, when Simpson's Rule or other rules should be used.

C provides a measure of the following and by considering the case when $V - v = \exp(-kt)$ (here $C = 1/(2k)$) it can be seen that $2C$ gives the "time-constant of the following."

§3. C-TEST FOR NON-UNIT FUNCTION INPUT.

3.1. Input Expressible in Operational Calculus Form.

Let the input

$$V(t) = \frac{h(p)}{k(p)} \cdot I \quad . \quad . \quad . \quad . \quad . \quad (3.1.1)$$

where $h(p)$, $k(p)$ are polynomials and I is the unit function ; we also make the proviso that $k(0) \neq 0$. Then

$$v(t) = \frac{A(p)}{B(p)} \cdot \frac{h(p)}{k(p)} \cdot I \quad . \quad . \quad . \quad . \quad . \quad (3.1.2)$$

and

$$\begin{aligned} V(t) - v(t) &= \frac{h(p)}{k(p)} \cdot p \cdot \frac{F(p)}{B(p)} \cdot I = p \frac{F'(p)}{B'(p)} I, \text{ say} \\ &= p \cdot \frac{f'_0 + f'_1 p + \dots}{b'_0 + b'_1 p + \dots} \cdot I \quad . \quad . \quad . \quad . \quad . \quad (3.1.3) \end{aligned}$$

Formula (2.1) for C again applies with the f 's and b 's replaced by the dashed letters.

If $k(p)$ contains a factor p , more consideration is needed, since, in this case, the difference $(V - v)$ will not approach zero as $t \rightarrow \infty$ and C will be infinite. Considering a special case $h(p)/k(p) = 1/p$ we see that, here, $(V - v) \rightarrow (b_1 - a_1)$ as $t \rightarrow \infty$ and either b_1 must be made equal to a_1 or, if a finite difference is to be allowed, in estimating the "goodness" of following we must consider $V - (b_1 - a_1) - v$. Other cases must be treated similarly.

* First suggested by Professor Wiener (1942), see James, Nichols and Phillips (1947).

3.2. Input Expressible as Fourier Integral.

The input may be easily expressed in this form but not in the form of 3.1 (e. g. the square wave $V=\text{constant}$ from $t=0$ to T and zero elsewhere).

Suppose

$$V(t)=\int_{-\infty}^{\infty} [H(w)+iK(w)] \exp (iwt) dw,$$

then

$$V(t)-v(t)=\int_{-\infty}^{\infty} [H(w)+iK(w)] iw \frac{F(iw)}{B(iw)} \exp (iwt) dw,$$

and by Parseval's Theorem,

$$\int_{-\infty}^{\infty} [V(t)-v(t)]^2 dt=(1/2\pi) \int_{-\infty}^{\infty} [\{H(w)\}^2 + \{K(w)\}^2] w^2 \frac{[f(w^2)]^2 + w^2 [g(w^2)]^2}{[b(w^2)]^2 + w^2 [c(w^2)]^2} dw$$

in the notation of 1.3.

§ 4. FURTHER PROCEDURE.

The case when C is a minimum has usually a pronounced overshoot (see fig. 1) and is usually considered undesirable. There are several ways of proceeding further to obtain cases with less overshoot.

4.1. The D-Test.

This is applicable when

$$V=\{h(p)I\}/k(p).$$

We define D as

$$\int_0^{\infty} [V-v] dt.$$

It is shown in the appendix 8.4 that

$$D=[h(o).f_0]/[k(o).b_0],$$

provided $k(o) \neq 0$. Now by choosing cases when C is near the minimum (within 6 or 7 per cent) but D is as positive as possible, we cut down the overshoot without losing much rapidity in following.

If, however, $f_0=0$ or $h(o)=0$ whatever the parametric values, then the D -test gives no help.

D may sometimes be found fairly easily when V is only conveniently expressed as a Fourier Integral (e. g. when V is a square wave, though $D \equiv 0$ in this case).

4.2. The E-Test.

E is defined as

$$\int_0^{\infty} [t(V-v)]^2 dt.$$

It is proved in 8.2 that, writing

$$\frac{\partial}{\partial p} \left[\frac{F(p)h(p)}{B(p)k(p)} \right] = \frac{F''(p)}{B''(p)} = \frac{f_0'' + f_1''p + \dots}{b_0'' + b_1''p + \dots}, \text{ say,}$$

$$\text{then } \pi E = \int_0^\infty \frac{[f_0'' - f_2'' w^2 + \dots]^2 + w^2 [f_1'' - f_3'' w^2 + \dots]^2}{[b_0'' - b_2'' w^2 + \dots]^2 + w^2 [b_1'' - b_3'' w^2 + \dots]^2} dw. \quad (4.2.1)$$

E has this advantage over C that it weights a difference occurring later compared with an earlier difference of the same magnitude. It is more troublesome to evaluate as it takes about twice the number of computations.

E is not easily computed if V is only expressible as a Fourier Integral, in fact

$$E = \int_{-\infty}^{\infty} [t\{V(t) - v(t)\}]^2 dt = (1/2\pi) \int_{-\infty}^{\infty} \left| \frac{\partial}{\partial w} \left[\{H(w) + iK(w)\} \frac{iwF(iw)}{B(iw)} \right] \right|^2 dw. \quad (4.2.2)$$

§ 5. COMPARISON OF THE TESTS IN A SPECIAL CASE.

Let the input V be a Heaviside Unit Function and purpose

$$v = \frac{b^2}{p^2 + 2ap + b^2} V, \quad (5.1)$$

where b^2 is fixed and a may take all positive values. We wish to find the value of a giving optimum "following".

In the above case

$$C = (4a^2 + b^2)/(4ab^2)$$

and

$$E = [16a^6 - 4a^4b^2 + a^2b^4 + b^6]/(16a^3b^6).$$

Hence C has a minimum when $a = b/2$ and E when $a = 2b/3$. The output v , in these two cases, has been plotted in fig. 1, together with the critically damped case $a = b$. The "following" is definitely better in the first two cases but, when $a = b/2$, the "overshoot" is 16 per cent, while, when $a = 2b/3$, it is only 6 per cent and, in general, this case would appear the best.

The values of C for $a = b/2$, $a = 2b/3$, $a = b$ are, respectively, $1/b$, $25/(24b)$ and $5/(4b)$ while the corresponding values for D are $1/b$, $4/3b$ and $2/b$.

§ 6. STABILITY TEST.

It is desirable to be certain that the system is stable, since, amongst other things, the formula for C is obtained on this assumption. The graphical test given here has these advantages over Routh's criteria (1860) (i) it needs practically no new calculation, (ii) in any case above the 5th degree Routh's criteria need lengthy and careful computation, and (iii) if the system is unstable it is much easier to see which parameter to alter to correct this.

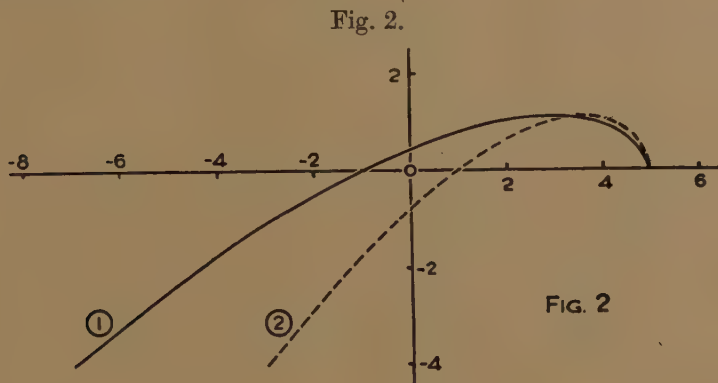
Test.—Plot $B(iw) \equiv b(w)^2 + iwc(w^2)$ (see (2.2)) on an Argand diagram. Treating this curve as the end of a vector the other end of which is attached to the origin we obtain the following rule:—

The system is stable (*i. e.* $B(p) = 0$ has no roots with positive real parts), if, and only if, this vector describes a rotation $n\pi/2$ anticlockwise as w goes from 0 to ∞ .

Fig. 2 shows two cases :

$$(1) B(p)=p^3+3p^2+2p+5 \text{ and } (2) B(p)=p^3+2p^2+2p+5.$$

In (1) the vector describes $3\pi/2$ anticlockwise and therefore indicates stability but in (2) it describes $\pi/2$ clockwise and indicates instability (proved in 8.3).



§ 7. FREQUENCY RESPONSE.

This is often required and can quickly be computed from data available from the calculation of C.

The frequency response is defined as the output when the input is sinusoidal, of unit amplitude, and of frequency $w/(2\pi)$ where w goes from 0 to ∞ . The output is given in amplitude and phase by

$$\begin{aligned} v(iw) &= A(iw)/B(iw) \\ &= \frac{b(w^2) + w^2g(w^2) + iw[c(w^2) - f(w^2)]}{b(w^2) + iwc(w^2)} \end{aligned}$$

in the notation of (2.2).

§ 8. APPENDIX.

8.1. Proof of Formula (2.2) for C.

From (1.2) when V is the Heaviside Unit Function I,

$$V-v = \frac{pF(p)}{B(p)} I = \frac{pA_r}{p+p_r} = A_r \exp(-p_r t) I$$

where $-p_r$ are the roots of $B(p)=0$.

Now $(V-v)$ can be expressed as a Fourier Integral, thus

$$\sqrt{(2\pi)} \cdot (V-v) = \int_{-\infty}^{\infty} g(w) \exp(iwt) dt,$$

where

$$\sqrt{(2\pi)} \cdot g(w) = \int_{-\infty}^{\infty} (V-v) \exp(-iwt) dt = \int_0^{\infty} (V-v) \exp(-iwt) dt$$

since $V=v=0$ when $t<0$. Hence

$$\begin{aligned}\sqrt{(2\pi)}g(w) &= \int_0^\infty [\Sigma A_r \exp(-p_r t)] \exp(-iwt) dt \\ &= \Sigma \frac{A_r}{p_r + iw} = \frac{F(iw)}{B(iw)}.\end{aligned}$$

By Parseval's Theorem

$$C \equiv \int_0^\infty [V-v]^2 dt = 2\pi \int_{-\infty}^\infty |g(w)|^2 dw.$$

Hence

$$\pi C = \int_{-\infty}^\infty \frac{F(iw)}{B(iw)} \frac{F(-iw)}{B(-iw)} dw$$

which on expansion gives (2.2).

8.2. *Proof of Formula (4.2.1) for E.*

Now

$$\begin{aligned}t(V-v) &= \Sigma A_r t \exp(-p_r t) = \Sigma \frac{p A_r}{(p+p_r)^2} \cdot I \\ &= p \frac{\partial}{\partial p} \left[\frac{F(p)}{B(p)} \right] \cdot I. \quad \dots \dots \dots (8.2.1)\end{aligned}$$

Applying the results of 8.1 to (8.2.1), we obtain formula (4.2.1) for E.

8.3. *Proof of Stability Test, §6.*

If the roots of $B(p)=0$ all lie to the left of the real axis then, by Cauchy's theorem the change of phase of $B(z)$ round a contour lying to the right of the real axis (or partly on it) must be zero. For our contour we take a semicircle of radius R and, as diameter, the imaginary axis from $+iR$ to $-iR$. For large R the change in phase of $B(z)$ in traversing the semicircle is practically that of the largest term $b_n z^n$, i. e. practically $n\pi$. Hence the change in phase in passing from $+iR$ to $-iR$ along the axis must be $-n\pi$. This proves the test of §6.

8.4. *Proof of Formula (4.1) for D.*

The operator $1/p$ is equivalent to $\int_0^t \dots dt$. It is proved in the literature (Berg 1929) that when p is put $=0$ in an operational formula, we obtain the result of putting $t=\infty$ in the equivalent formula. Applying this we obtain the formula given for D in (4.1).

REFERENCES.

- BERG, E. J., 1929, *Heaviside's Operational Calculus* (1936, McGraw-Hill), 2nd ed. Chap. III.
JAMES, H. M., NICHOLS, N. B., and PHILLIPS, R. S., 1947, *Theory of Servo-Mechanisms* (Massachusetts Institute of Technology Publications).
PORTER, A., and MACK, C., 1949, "New Methods for the Numerical Solution of Algebraic Equations", *Phil. Mag.* [7], **40**, 578.
ROUTH, E. J., 1860, *A treatise on the dynamics of Rigid Bodies* (1905, Macmillan), Part II, p. 170, 6th ed.

LXXXVI. *Investigation of Soft Radiation by Proportional Counters—V.
Use as a Detector of Ultra-violet Quanta and Analysis of the Gas
Multiplication Process.*

By S. C. CURRAN, A. L. COCKROFT, and J. ANGUS,
Department of Natural Philosophy, University of Glasgow*.

[Received July 11, 1949.]

SUMMARY.

Quantitative detection of single slow electrons liberated by light at the cathode of a proportional counter has been observed and good plateaux obtained. The distribution in amplitude of the output pulses has been measured and the shape of this distribution has been shown to correspond rather closely with the curve $x^{\frac{1}{2}} \exp(-x)$. The variation introduced by the gas gain alone has been calculated for various fixed amounts of ionization and compared with the experimental observations on homogeneous radiations. By this means the fluctuations in the number of ion pairs released in the gas by electrons totally absorbed in the counter has been deduced. It appears that the total variation can thus be split into its two main components. These are found to be of about equal magnitude and less than predicted by current theory.

INTRODUCTION.

It has been shown in work previously reported (Curran *et. al.* 1949) that the proportional counter operating in conjunction with an amplifier of high gain and good signal-to-noise performance can be used to measure accurately the energy of soft ionizing radiations. Attention was drawn to the fact that the same arrangement could be used to measure accurately the total number of low-energy particles, *e.g.* β -rays of H^3 or soft X-ray quanta which produced ionization within the tube. In particular, the proportional counter can be regarded as a detector of 100 per cent efficiency for such radiations, possessing long plateaux of less slope than are usually obtained with Geiger tubes. This fact is important for several reasons among which may be noted the increased life of the proportional counter as compared with the life of a vapour-gas Geiger tube (factor 10^4 to 10^5) and the relative freedom from after-pulses produced by bombardment of the cathode by positive ions liberated in the discharges.

It has been established that a proportional counter can operate satisfactorily at gas multiplications of 10^4 or perhaps more. Thus with an amplifier capable of detecting down to, say, 2×10^3 electrons at the

* Communicated by the Authors.

first grid, it should be possible to count single electrons liberated in the counter. Moreover, if the gas gain can be made sufficiently large without impairing the stability of operation, the analysis of the energy spectrum of the pulses produced by single electrons can be carried through. Thus fundamental data concerning the nature of the statistical fluctuation of gas amplification can be obtained. The use of single electrons as the primary particles permits the separate study of this fluctuation in the absence of the complicating factors introduced by the variations in the number of ion pairs initially produced when homogeneous β -rays, X-rays or γ -rays are introduced. The study of the mode of operation of a proportional counter when detecting single photoelectrons was therefore regarded as of primary importance, since it seemed to offer (i) the possibility of realizing an efficient and reliable detector of ultra-violet quanta and (ii) the necessary data for clarification of the statistics of the multiplication process. Such information is important on account of its basic relation to the resolving power of the proportional tube as an analyser of radiation.

In this work, the proportional counter has been found to operate very satisfactorily as a detector of ultra-violet radiation and a study of the pulse size distribution for single electrons has enabled us to separate the contribution to line-width due to the gas multiplication process and the variations in the initial ionization produced by homogeneous radiation.

DETECTION OF ULTRA-VIOLET.

Several counters were used as photon detectors in the course of the experiments, but the results were identical in all cases. We show in fig. 1 the simple form of tube with which most of the measurements were made. The cathodes consisted of semi-transparent layers of aluminium, which were evaporated on to the quartz envelope forming the body of the counter. Almost equally good performance was secured more easily by using cathodes of polished aluminium foil and by mounting a thin window (~ 0.2 mm.) of quartz over an aperture in the cylinder (see fig. 1). The mixture within the counter consisted of argon and methane, each to a pressure of 25 cm. of Hg. The efficiency of the cathode seemed to be $\sim 10^{-3}$ photoelectrons per quantum at wavelengths in the region of 3000 to 3500 angstroms and a satisfactory source of radiation was secured by operating a small tungsten lamp on an accumulator.

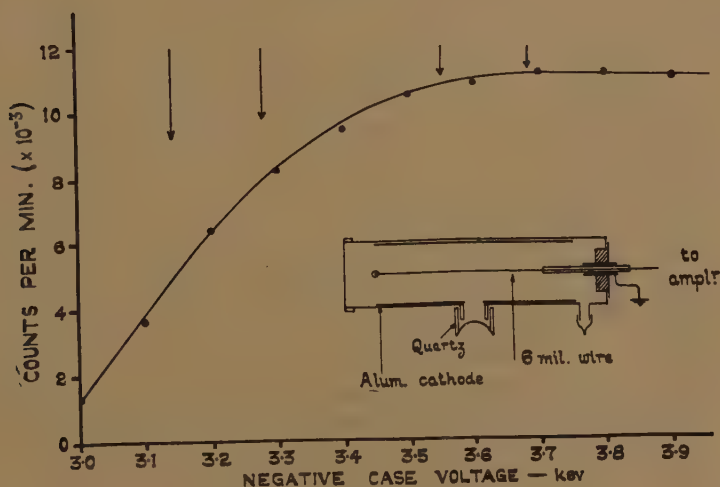
The counter started to operate at a little over 2000 volts, and at 2300 volts the L X-rays of bismuth from a standard source of radium D+E+F (covered with polythene to remove the β -rays) gave output signals ~ 50 volts amplitude. The average amplitude of the background pulses due to cosmic radiation, etc., was of the same order at this level; it is in this region of gas gain that most of our earlier studies of the performance of the tubes were carried out.

As the voltage was increased, signals due to the source of light began to stand out above the noise peaks and at about 3000 volts a large fraction

were of sufficient magnitude to operate the scaling unit (threshold sensitivity 3 volts). The rate of counting increased smoothly as the voltage was raised; the results above 3 kv. are shown in the curve of fig. 1. The background pulses, as distinct from those due to light, were enormously greater in amplitude and saturated the amplifier over the whole of this range. A plateau on the counting curve was obtained between about 3.5 and 3.9 kv.; between 3.7 and 3.9 kv. the slope was effectively zero. In this region, therefore, the proportional counter forms a sensitive and quantitative detector of light.

It was not possible to establish the proportionality of the counter in the vicinity of 3.9 kv. by any of the methods previously adopted (Curran *et al.* 1949), *e.g.* by passing soft homogeneous X-radiation of known

Fig. 1.

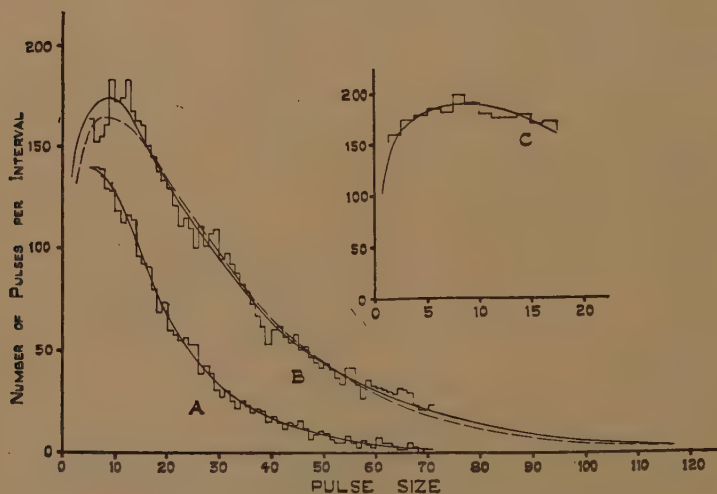


Counting Plateau for single electrons in proportional counter.

quantum energies into the tube. It was therefore necessary to establish proportionality by the observations made on the pulse distribution arising from single electrons released at the wall by ultra-violet quanta; the shape of the spectrum at different voltages offers a rather sensitive method of test. The spectrum was examined at 3145, 3280, 3550 and 3685 volts. It was found that the four spectra obtained could be fitted closely to each other; two examples (at 3280 and 3550 volts) are shown in fig. 2. Any appreciable departure from proportionality, which normally becomes evident as the Geiger threshold is approached, would have been indicated as a convergence of the pulses to a more uniform value as the voltage was increased. No such tendency is evident up to at least 3685 volts, which point is in the middle of the plateau for single electrons. Although we did not carry through a complete analysis of the distribution at 3.9 kv., we believe that the tube behaved proportionally throughout the range of voltage of fig. 1. The points examined in detail are marked by arrows in fig. 1.

Calculation of the gain independently confirms the proportionality of operation at large multiplication values. At 2340 volts the amplitude of the pulses due to the L X-radiation (Bi) emitted by radium D of energy 13 kev. was noted. At 3550 volts the average pulse size (measured from the spectrum in fig. 2) due to single electrons was found to be three times larger. Assuming that the energy required to produce one ion pair is ~ 30 ev., the X-radiation gives rise to $13,000/30$, *i.e.* 433 ion pairs in the tube, and the gas amplification has increased by a factor of 3×433 , *i.e.* 1300, for an increase of 1210 volts. The rates of increase of gas gain with voltage were measured at 2340 volts and 3550 volts and found to correspond to a doubling of the pulse size for increments of 135 and 105

Fig. 2.



Pulse Size Distribution for single electrons in proportional counter.

volts respectively. In the first case the rate was measured by observations on the homogeneous group of pulses due to the X-rays of Radium D and in the second case by measuring the average pulse size of the spectrum. A multiplication of the pulse size by 1300 for an overall increase of 1210 volts corresponds to a doubling for each rise of 118 volts. This rate lies between the extreme values of 135 and 105 actually measured, and it would therefore appear that the rate of increase of gain with voltage rises smoothly and relatively steadily over the very wide range of gain involved here.

It is of interest to calculate the total tube gain at 3550 volts. The L X-ray line of bismuth which gives rise to 433 ion pairs was estimated very approximately to be about 20 times greater than the peak noise signals, which correspond approximately to the signal produced by the collection of about 2500 ion pairs at the first grid. Hence at 2340 volts the gas gain was $20 \times 2500/433$, or 113, and hence the total gas amplification

at 3550 volts was 113×1300 , or 1.46×10^5 . It is apparent that multiplication in excess of 10^5 can be obtained in proportional tubes provided that the initial amount of ionization is very small. For most radiation 10^4 is accepted as an approximate upper limit, but this figure is conditioned entirely by the ionization level. At 3.9 kv. (fig. 1) the multiplication is greater than 10^6 .

STATISTICAL ANALYSIS.

In fig. 2 curves A and B give the pulse spectra as measured at 3280 and 3550 volts respectively; curve C shows part of curve B as analysed at higher amplifier gain. Curves A and C have been used to extend curve B and to give an overall picture of the single electron distribution for pulse sizes between 2 and 120 units. The full curve through these points has a maximum at 9 units and is considerably asymmetric about this value. The mean value of the curve (\bar{A}) and the root mean square deviation (σ_A) were evaluated as 27.1 units (corresponding to a multiplication of 1.46×10^5) and 22.6 units respectively. It is considered from theoretical studies of the gas multiplication process that σ_A should be directly proportional to \bar{A} over a large range of voltage, and this assumption appears to be borne out by the experimental results; the value of σ_A^2/\bar{A}^2 calculated from curve B is 0.696.

An attempt was made to fit a curve of the form $x^{(n-1)} \cdot \exp(-x)$ to this distribution, since this equation has relatively simple statistical additive properties. A close fit was obtained with $n=3/2$ when the mean values of the two curves were made equal; this curve (broken) is also given in fig. 2. The mean value and the root mean square deviation of this distribution may be evaluated as follows:—

$$\begin{aligned}\bar{x} &= \int_0^\infty x^{3/2} \exp(-x) dx \bigg/ \int_0^\infty x^{1/2} \exp(-x) dx \\ &= \Gamma(5/2)/\Gamma(3/2) = 3/2, \\ x^2 &= \int_0^\infty x^{5/2} \exp(-x) dx \bigg/ \int_0^\infty x^{1/2} \exp(-x) dx \\ &= \Gamma(7/2)/\Gamma(3/2) = 15/4, \\ \sigma_x^2 &= \overline{x^2} - (\bar{x})^2 = 3/2.\end{aligned}$$

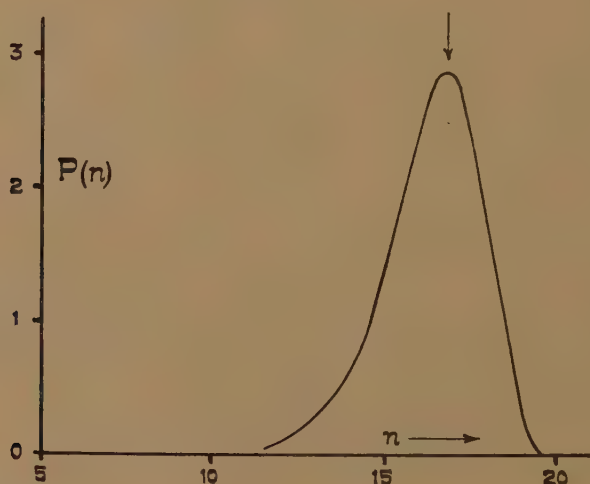
The ratio $\sigma_x^2/(\bar{x})^2$ is therefore equal to 0.666 and this value is rather close to the value of σ_A^2/\bar{A}^2 calculated from the experimental distribution (0.696); throughout the subsequent work the average value of 0.681 has been used for this ratio.

On the assumption that the gas gain A is given by 2^n , where n is the number of ionizing collisions made by the photoelectrons on their passage to the collecting wire, a probability distribution in terms of n may be

derived from curve B. This distribution is given in fig. 3; it shows a maximum at $n=17$, which corresponds almost exactly with the value derived from the mean value of A according to the equation $2^n = 1.46 \times 10^5$. The curve is slightly asymmetric about the maximum, the mean value of n being 16.2; the distribution is surprisingly narrow, and the root mean square deviation, 1.4, gives evidence for a considerable uniformity in the number of collisions made. The ratio σ_n/\bar{n} (0.087) is the same as that which would be obtained on the assumption that the gas multiplication is equal to $\exp(n)$, where n is the number of mean free paths for ionizing collisions.

With the fitted curve $x^{1/2} \exp(-x)$, it is possible to calculate the shape of the pulse distribution which would be obtained from a homogeneous group of N electrons. The shape of this distribution is given by $x^{(Nm-1)} \exp(-x)$, where m , as before, is equal to $3/2$; the mean value is

Fig. 3.



Variation of ionizing collisions in gas multiplication process.

Nm and the root mean square deviation is $(Nm)^{1/2}$. The variation in pulse-size due to the multiplication process alone may therefore be written as

$$\sigma_P^2/\bar{P}^2 = 1/N \times \sigma_A^2/\bar{A}^2 \quad 0.681/N,$$

where \bar{P} is the average pulse size and σ_P is the root mean square deviation of this size.

The variations in the loss of energy of a primary electron by ionization collisions impose a fluctuation in the number of ion pairs produced by electrons of equal energy. For high energies this fluctuation is assumed to be equal to the square root of the average number of ion pairs formed, but there is no reason to suppose that this relation should be valid for slower electrons which dissipate all of their energy in the gas through which they travel. Suppose then that an electron produces N_r ion pairs.

along its complete track; the mean value of $N_r(\bar{N})$, will be equal to E/E_i , where E is the energy of the electron and E_i is the average energy lost per collision. Fluctuations in N_r will arise mainly from spread in the value of the energy loss per ionizing collision.

The root mean square deviation in the number of electrons produced may be defined by the equation

$$\sigma_N^2 = 1/R \sum_1^R (N_r - \bar{N})^2, \quad R \rightarrow \infty.$$

Assuming that each negative ion (electron) produced is amplified independently in the multiplication process, an assumption which should certainly be valid for the relatively small values of N under discussion, the pulse size obtained from N_r electrons may be written as

$$P_{N_r} = \sum_1^{N_r} A_j,$$

and the following relations may be obtained

$$\bar{P} = \bar{A} \cdot \bar{N}$$

and

$$\sigma_P^2/\bar{P}^2 = \sigma_N^2/\bar{N}^2 + 1/\bar{N} \times \sigma_A^2/\bar{A}^2. \quad (1)$$

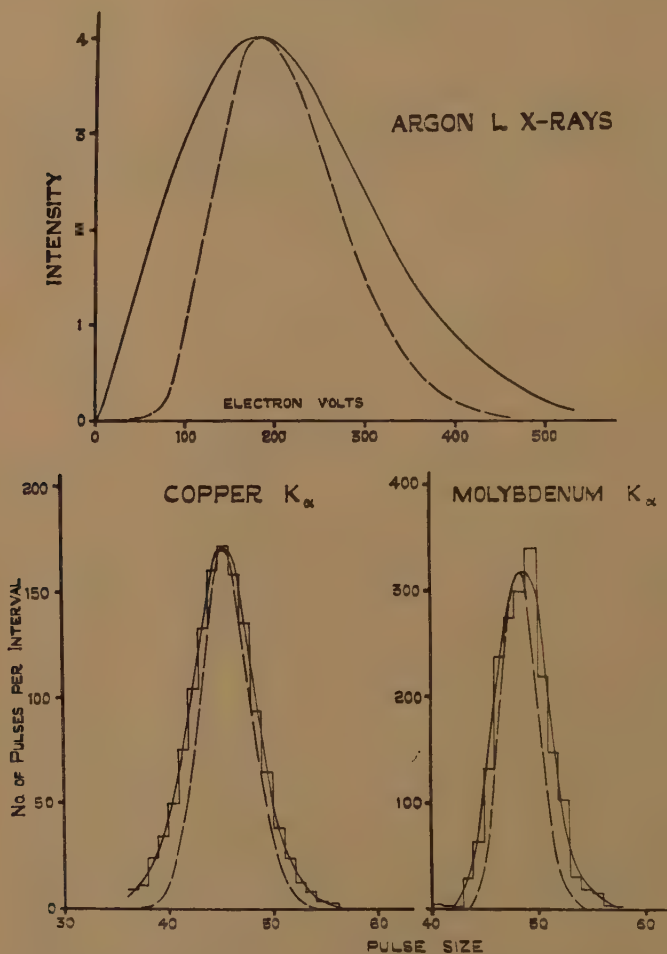
We have carried out more detailed experiments on the K_α X-rays of copper and molybdenum, using selectively absorbing filters to remove the K_β groups; from these distributions and from the distribution of pulses for the L X-rays of argon given by Hanna *et al.* (1949), we have calculated the ratio σ_P/\bar{P} . The distributions obtained are given in fig. 4 together with the distributions calculated from the relation $x^{(2/3 \cdot \bar{N}-1)} \exp(-x)$, which represent the fluctuations due to gas multiplication alone. The latter are shown by the broken curves. For the calculation of \bar{N} , we have assumed that the average energy loss E_i is of 32 electron volts per ion pair, and with this value we have calculated the fluctuation introduced by the gas multiplication process. Hence, by using equation (1), we have determined the value of σ_N . The assumption of an average value of 32 ev. for slow electrons is rather tentative in view of the wide range of values for argon given by the measurements of different experimenters and since the value for methane is not well established. Our results depend fairly critically on the actual value adopted, and we hope to measure E_i for argon and methane in the near future. The results of the calculations are given in the following table:—

Energy (kev.)	\bar{N}	$\sigma_P/\bar{P}\%$	$(\bar{N})^{-1/2} \sigma_A/\bar{A}\%$	$\sigma_N/\bar{N}\%$	σ_N	$\sigma_N/(\bar{N})^{1/2}$
0.19	6	50.25	33.7	37.2	2.24	0.91
8.0	250	7.43	5.21	5.29	13.2	0.84
17.4	550	5.00	3.53	3.53	19.4	0.83

CONCLUSIONS.

It is apparent from the analysis of the pulse distribution that, over most of the normal range of operation of a proportional tube, the contributions to line width arising from fluctuations in the initial number of ion pairs and the variations in the gas multiplications are approximately equal.

Fig. 4.



Analysis of line-width for mono-kinetic radiations.

over a wide range of energy. The total fluctuation is shown, at least for a mixture of equal quantities of argon and methane, to be given approximately by

$$\sigma_P^2/\bar{P}^2 = (\alpha + \beta)/\bar{N},$$

where $\alpha = 2/3(0.681)$ and $\beta = 2/3$ for $E_i = 32$ ev.; this variation is smaller than that given by Snyder (1947) and may indicate the need for a modified form of the theory. With the known value of α , viz. $2/3$, β would have to

be $4/3$, a value in considerable disagreement with our result and one which would require an unexpectedly low value for the energy required to produce an ion pair in the gas mixture involved here. As an indication of this dependence, the assumption of a value of 27 ev. leads to a value of $\beta=1$.

It is hoped that these investigations can be extended to include measurements on a number of gases and mixtures of gases at different pressures and to examine the effect of different tube geometries of the resolution

We have pleasure in thanking Professor P. I. Dee for continual help and advice throughout the course of the work.

REFERENCES.

- CURRAN, S. C., ANGUS, J., and COCKROFT, A. L., 1949, *Phil. Mag.*, **40**, 36, and 53.
HANNA, G. C., KIRKWOOD, D. H. W., and PONTECORVO, B., 1949, *Phys. Rev.*, **75**, 985.
SNYDER, H. S., 1947, *Phys. Rev.*, **72**, 181A.

LXXXVII. *Non-symmetric Stress-Energy-Momentum Tensor and Spin-Density.*

By A. PAPAPETROU,

I.C.I. Research Fellow, University of Manchester *.

[Received June 14, 1949.]

ABSTRACT.

The question of the physical meaning of a non-symmetric stress-energy-momentum tensor is discussed. It is shown that an antisymmetric part of this tensor arises necessarily if there are particles with spin which must be considered as *point* particles. Some formulæ connecting the antisymmetric part of T_{ik} and the spin density are derived. It is shown that the unified field theory of Einstein and Straus with non-symmetric metrical tensor does not lead to an explanation of geomagnetism.

§1.

A NON-SYMMETRIC stress-energy-momentum tensor has been already considered in relativistic field theories, where it arises automatically from the general formalism which leads to the identities representing the conservation laws. But this formalism cannot give an answer to the question whether or not the antisymmetric part T_{ik}^{\vee} of T_{ik} has any direct *physical meaning*. Indeed, this antisymmetric tensor can be modified arbitrarily by adding terms of the form of a divergence to T_{ik} ; and it can be made even zero by the general procedure for the symmetrization of T_{ik} .

* Communicated by Professor Rosenfeld.

However, it is clear from the very beginning that an antisymmetric tensor T_{ik} can have some physical meaning only if the field which we are considering contains *particles with spin*. The simplest way to arrive at this conclusion is to consider the formulation of the angular momentum law. When T_{ik} is non-symmetric, the angular momentum density contains not only the usual two terms involving T_{ik} , but also a third term which then has to be interpreted as representing a spin-density. A much more direct and elementary argument for this point of view is the following.

In special relativity we usually require the tensor T_{ik} to be symmetric, this being the generalization of the symmetry of the stress-tensor $t_{\alpha\beta}$ ($\alpha, \beta = 1, 2, 3$) in Newtonian mechanics. If we analyse the last condition, we see that it is based on the assumption that in the material system which we are considering there is only *orbital* angular momentum. We can then argue that the forces acting upon a volume element dV , whose linear dimensions are of the order of magnitude dl , will possess a moment about a line passing through the element of the order of magnitude $(t_{\alpha\beta} - t_{\beta\alpha})dl^3$; on the other hand, the angular momentum of the material inside the volume element about this line is proportional to $dV \cdot dl^2 \sim dl^5$, and consequently, by equating the moment of the forces to the rate of change of angular momentum, we conclude that $t_{\alpha\beta}$ must be symmetric. Now we know that the elementary particles have a spin, and that there are various effects resulting in a partial orientation of the spins in some fixed direction (*e. g.* gyromagnetic effect). When such an effect is present, the total angular momentum inside the volume element dV will contain an additional term representing the sum of the oriented spins; this term being proportional to $dV \sim dl^3$, we see that the previous argument for the symmetry of $t_{\alpha\beta}$ will not hold any more.

In the following sections the question of the physical meaning of T_{ik} will be discussed in some detail. In order to avoid unnecessary complications, the simplest case of a "perfect gas" of spinning particles (*i. e.* a system of non-interacting spinning particles) will be treated. The considerations will be based on general principles of mechanics and will lead us to some formulæ for T_{ik} . The main conclusions to which we shall arrive may be stated here: T_{ik} will be *necessarily non-symmetric* if the spinning particles must be considered as *point particles* (*i. e.* if it would have no physical meaning to speak of an internal structure of the particles). The calculations will be classical throughout, but the final formulæ can be transcribed for quantum-mechanical systems by the use of the general principle of correspondence.

The author has been interested in this problem for the following reason. He was trying to discuss the possibility of a general explanation of geomagnetism, as proposed by Blackett (1947), in the frame of a unified theory of gravitation and geomagnetism. While such an explanation has been impossible in all the theories based on a symmetric metrical tensor g_{ik} , it seemed interesting to discuss the problem in the frame of the theory with

non-symmetric g_{ik} which has been proposed recently by Einstein and Straus (1946). A detailed discussion, necessarily connected with the consideration of a non-symmetric tensor T_{ik} , has been made by the present author, but the result has been negative. A short account of the assumptions on which these calculations have been based and of the final results will be given in the last section.

§ 2.

We have already mentioned in the previous section that we shall derive formulæ for T_{ik} by means of entirely classical calculations. In order to justify the use of such calculations, it will in the first instance be sufficient to remark that it is possible to describe a "point" particle—i. e. a particle with linear dimensions tending to zero—having an intrinsic angular momentum by using the conceptions of classical relativistic mechanics only. The unusual feature of such a particle is that it cannot be described by a mass pole only but also dipole and more generally higher multipole elements will be necessary, as was first suggested by Mathisson (1937).

The simplest type of a pole-dipole particle has been discussed in detail (Hönl and Papapetrou 1939 a, 1939 b). In the system of coordinates in which the momentum of this particle vanishes, the particle can be described as follows. It is a "point" particle moving on a circle of a given radius with a constant velocity having any value smaller or equal to c . The centre of the circle coincides with the mass centre of the particle: The mass centre lies outside the particle because of its pole-dipole structure. The angular momentum of the particle, coinciding in this frame (because of the vanishing momentum) with its spin, is normal to the plane of the orbit.

It is now remarkable that, as shown by a detailed comparison (Hönl and Papapetrou 1940), this pole-dipole particle has a dynamical behaviour which is surprisingly similar to the behaviour of the electron according to Dirac's theory; so far going is the similarity that it seems justified to consider this particle as the *classical model* of the Dirac electron. This result makes it almost certain that at least an essential part of the conception of spin can be described in terms of classical mechanics. We thus have an additional justification for the use of classical calculations for spinning particles which we are going to make in the following sections.

It will be useful to recall the difference in the dynamical description of a spinning particle as against a non-spinning single-pole particle. The latter is fully described by its rest mass m_0 and the four dimensional velocity $u^i = dx^i/ds$. The momentum is then given by

$$p^i = m_0 u^i,$$

while an intrinsic angular momentum does not exist. In the case of a spinning particle there is no longer any direct relation between momentum and velocity, as we see from the description of the pole-dipole particle given above: The (3-dimensional) momentum \vec{p} of the particle was zero, while its velocity \vec{v} did not vanish. Therefore a spinning particle will be

The differential form of the angular momentum law is the following:

$$\frac{\partial}{\partial x^s} F^{iks} = \frac{\partial}{\partial x^s} (l^i T^{ks} - l^k T^{is} + f^{iks}) = 0, \quad \quad (10)$$

f^{iks} representing the spin contribution to the total angular momentum density F^{iks} . Writing (10) in detail, *e. g.* for $i=1, k=2$,

$$\frac{\partial F^{121}}{\partial x} + \frac{\partial F^{122}}{\partial y} + \frac{\partial F^{123}}{\partial z} + \frac{\partial F^{124}}{c \partial t} = 0,$$

and remembering that F^{124} is the density of angular momentum "parallel to the hyperplane $x^1 x^2$ " (in Newtonian mechanics parallel to the axis Oz) we see that F^{12s} must be considered as the flux of the component $x^1 x^2$ of angular momentum through a unit surface normal to Ox^s ; and generally F^{iks} will be the flux of the component $x^i x^k$ of angular momentum through a unit surface normal to Ox^s . From the differential law (10) we derive the integral conservation law (Papapetrou 1939)

$$c J^{ik} = \int F^{ik4} dv = \text{const.}, \quad \quad (11)$$

the six quantities J^{ik} being the components of an antisymmetric tensor.

We now consider the angular momentum law for one particle of our gas. We denote by j^{ik} the (constant) angular momentum of this particle for the centre of momenta ξ^i and by j_0^{ik} its intrinsic angular momentum, *i. e.* the value which we find if we take as centre of momenta the position X^i of the particle at the corresponding time. A relation between j^{ik} and j_0^{ik} can be derived immediately if we make use of the general transformation formula for a change $\xi^i \rightarrow \xi^{*i}$ of the centre of momenta which reads as follows (Papapetrou 1939):

$$J^{ik} = J^{*ik} + (\xi^{*i} - \xi^i) P^k - (\xi^{*k} - \xi^k) P^i,$$

P^i being the energy-momentum fourvector

$$c P^i = \int T^{i4} dv.$$

"Taking $\xi^{*i} = X^i$ and writing j^{ik} , j_0^{ik} and p^i instead of J^{ik} , J^{*ik} and P^i we find

$$j^{ik} = j_0^{ik} + (X^i - \xi^i) p^k - (X^k - \xi^k) p^i. \quad \quad (12)$$

Again we stress that the quantities j_0^{ik} will vanish identically only in the case of particles having the structure of a single pole; but in the case of a more complicated structure they will usually have non-vanishing values.

Assuming now that all the particles of the gas have the same values of j_0^{ik} and v^i and applying the arguments which lead us to (4) we find

$$F^{ikl} = n j_0^{ik} v^l; \quad \quad (13)$$

and if we take into account (12) and (4):

$$F^{ikl} = n j_0^{ik} v^l + l^i T^{kl} - l^k T^{il} \quad \quad (13 \alpha)$$

Comparing with (10) we see that

$$f^{ikl} = n j_0^{ik} v^l. \quad \quad (14)$$

The generalization for the case of particles having different values of j_0^{ik} and v^l is straight forward. We find

$$f^{ikl} = n \overline{j_0^{ik} v^l}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (14 a)$$

n being again the total density of particles and $\overline{j_0^{ik} v^l}$ the average of $j_0^{ik} v^l$ for all particles.

Taking account of (3) we find from the differential angular momentum law (10)

$$T^{ki} - T^{ik} + \frac{\partial}{\partial x^s} f^{iks} = 0,$$

i. e.

$$T_V^{ik} \equiv \frac{1}{2} (T^{ik} - T^{ki}) = \frac{1}{2} \frac{\partial f^{iks}}{\partial x^s} \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

This is the relation connecting the antisymmetric part of T^{ik} to the spin-density f^{ikl} . It is clear from the way we have derived it that equation (15) is valid generally and not only for a perfect gas. But we must not forget that for a general application of (15) we ought first to generalize equation (14). The expression (14) or (14 *a*) for f^{ikl} is valid for a perfect gas only.

We mention how equation (14 *a*) can be transcribed for Dirac's theory of the electron. The operator corresponding to j_0^{ik} is (Papapetrou 1940)

$$j_0^{ik} = \frac{i\hbar}{4} (\alpha_0 \alpha^i \alpha_0 \alpha^k - \alpha_0 \alpha^k \alpha_0 \alpha^i). \quad . \quad . \quad . \quad . \quad . \quad . \quad (16 a)$$

It is worth noticing that this formula too can be derived from the expression giving j_0^{ik} for the classical model of the electron described in §2 by using the principle of correspondence (Hönl and Papapetrou 1940). With (16 *a*) and (8 *a*) we find

$$f^{ikl} = -\frac{i\hbar c}{4} \psi^* \alpha^l (\alpha_0 \alpha^i \alpha_0 \alpha^k - \alpha_0 \alpha^k \alpha_0 \alpha^i) \psi, \quad . \quad . \quad . \quad . \quad . \quad . \quad (16 b)$$

which has been shown to be the correct expression for the spin-density in Dirac's theory (Papapetrou 1940).

§5.

If we think more carefully on the calculations made in the last two sections, we see that they are inseparably connected with the following assumption: A particle which is going to cross a surface element dS does it *at once*; *i. e.* at any given time we could only say that the particle is on the one or the other side of dS , but there would be no physical meaning in the statement that a part of the particle is on the one and the remaining part on the other side of dS . In other words, T_V^{ik} and f^{ikl} will have a direct physical meaning if the particles must be considered as *point particles*, this being equivalent to the statement that it has no physical sense to speak of an internal structure of these particles. It follows that an orientation of the angular momenta of molecules or atoms and even of compound nuclei must in any case be disregarded in the calculation of a physically meaningful

$T^{\bar{ik}}$ or f^{ikl} . The only possible contributions would come from the spins of the elementary particles (electrons, protons, neutrons); though it must be added that even this is doubtful, since it is not impossible that even for these particles it might be justified to speak of an intrinsic structure of the particles.

If each particle had an intrinsic structure describable by a *symmetric* tensor t^{ik} —as it seems entirely justified to assume, *e. g.* for molecules and atoms, but as is also the case for the classical model of the electron which we mentioned in § 2—, the tensors $T^{\bar{ik}}$ and f^{ikl} can still be defined by means of (4) and (14). But these quantities must then be considered as auxiliary quantities which might be convenient for special purposes. What would in this case have a direct physical meaning is the symmetric tensor \bar{t}^{ik} which arises if we average t^{ik} over a volume element dV whose linear dimensions are large compared with the mean distance of the particles. Calculations which we shall not repeat here lead to the relation

$$\bar{t}^{ik} = T^{ik} - \frac{1}{2} \frac{\partial}{\partial x^s} (f^{iks} + f^{sik} + f^{ski}), \quad (17)$$

T^{ik} and f^{iks} being defined by (4) and (14). We see that \bar{t}^{ik} is exactly the tensor arising from the general method of symmetrization of T^{ik} (Belinfante 1939, Rosenfeld 1940).

The last possibility which we ought to mention is that the particle might have an intrinsic structure describable in terms of a *non-symmetric* tensor t^{ik} . The procedure of averaging t^{ik} over the volume dV would then lead to a non-symmetric tensor \bar{t}^{ik} having a non-vanishing antisymmetric part $\bar{t}^{\bar{ik}}$. It is easy to see that $\bar{t}^{\bar{ik}}$ would be identical with the $T^{\bar{ik}}$ calculated from (4) or (15) if the symmetric part of the tensor t^{ik} describing the particle had the structure of a single pole (in which case this symmetric part cannot contribute to the spin of the particle); if this is not so, $\bar{t}^{\bar{ik}}$ should be calculated from the detailed structure of the particle.

§ 6.

In this last section we describe very briefly an (unsuccessful) attempt for an explanation of geomagnetism in the frame of the unified field theory with a non-symmetric g_{ik} . Such a discussion seemed to be worthwhile since there is in this case a very simple possibility of correlating the electromagnetic to the gravitational field: There are two vector densities, built up only from first derivatives of g_{ik} , which both fulfil the continuity equation and consequently are suitable for identification with the current-charge density. These are

$$s^i = \frac{\partial g^{\bar{ik}}}{\partial x^k}, \quad s^{*i} = \epsilon^{iklm} \frac{\partial}{\partial x_m} g^{\bar{ik}}, \quad (18)$$

ϵ^{iklm} being the totally antisymmetric tensor density of Levi-Civita.

Obviously this discussion had to be based on the assumption that the antisymmetric tensor $T^{\dot{ik}}$ has a direct physical meaning. The procedure has been the following. Taking into account the known effects of spin orientation inside the earth we calculate $T^{\dot{ik}}$. Then from the field equations for weak fields we derive the tensor $g_{ik}^{\dot{ik}}$. Finally we try to see whether by using as current-charge density one of the quantities (18) it is possible to derive from this $g_{ik}^{\dot{ik}}$ a dipole magnetic field for the rotating earth.

As field equations we take the equations of general relativity

$$R_{ik} = -\kappa(T_{ik} - \frac{1}{2}g_{ik}T), \quad . \quad . \quad . \quad . \quad . \quad (19)$$

with the only difference that g_{ik} will now be non-symmetric. We assume the field to be weak and consider only the first approximation :

$$g_{ik} = \gamma_{ik} + g'_{ik},$$

γ_{ik} being the metrical tensor of special relativity and g'_{ik} the first order term. Equation (19) then reduces to

$$\square g'_{ik} = -2\kappa(T_{ik} - \frac{1}{2}\gamma_{ik}T).$$

Since γ_{ik} is symmetric, the antisymmetric part of this equation is

$$\square g'^{\dot{ik}}_{\dot{v}} = -2\kappa T^{\dot{ik}}_{\dot{v}}, \quad . \quad . \quad . \quad . \quad . \quad (20)$$

the indices being raised in this approximation by means of $\gamma^{\dot{ik}}$.

For an approximate calculation of $T^{\dot{ik}}$ we consider only the contribution of the free electrons inside the rotating body (conductivity electrons inside the very probably metallic core of the earth). Therefore it is permissible to make use of the following relation which has been derived from Dirac's theory (Tetrode 1928) :

$$\frac{\partial T^{\dot{ik}}_{\dot{v}}}{\partial x^k} = 0.$$

Introducing this in (20) we find

$$\square \frac{\partial g'^{\dot{ik}}_{\dot{v}}}{\partial x^k} = 0.$$

Since this equation is valid in the whole of space-time (inside matter as well as in empty space), it follows that

$$\frac{\partial g^{\dot{ik}}_{\dot{v}}}{\partial z^k} \simeq \frac{\partial g'^{\dot{ik}}_{\dot{v}}}{\partial x^k} = 0, \quad . \quad . \quad . \quad . \quad . \quad (20 a)$$

i. e. we have exactly the theory of Einstein and Straus (1946). As the current-charge density we can now use the quantity \mathfrak{s}^{*i} only, since $\mathfrak{s}^i \equiv 0$ according to (20 a). This means that we must take the electromagnetic field $F^{ik} \sim g^{\dot{ik}}_{\dot{v}}$ with the following interpretation :

$$(F^{23}, F^{31}, F^{12}) = \vec{E}, \quad (F^{14}, F^{24}, F^{34}) = \vec{H}.$$

Finally equation (20) takes the form

$$\square F^{ik} = \lambda T^{ik}, \quad (21)$$

λ being a constant having the dimensions of a reciprocal charge*.

There are two known effects of spin orientation inside the earth. The first is the spin orientation due to the rotation of the earth; and the second the orientation of the magnetic moment of the electron—and consequently of its spin—by the actual magnetic field of the earth. Both effects tend to orient the spin inside the meridian plane. If we calculate the corresponding T^{ik} from (15) and then the field F^{ik} from (21), we find a magnetic field normal to the meridian plane (lines of force circles around the axis of rotation). Simple considerations show that in order to get a magnetic field in the meridian plane—as is the dipole magnetic field—we must start from an effect which tends to orient the spins in the direction of the normal to the meridian plane, *e. g.* by assuming an additional magnetic field normal to the meridian plane. But an explanation of this kind would be so complicated that we rather must conclude that the unified field theory with non-symmetric g_{ik} fails to explain the magnetic field of the earth.

The author wishes to thank Professor L. Rosenfeld for valuable discussions.

REFERENCES.

- BELINFANTE, F., 1939, *Physica*, **6**, 887.
 BLACKETT, P. M. S., 1947, *Nature, Lond.*, **159**, 658.
 EINSTEIN, A., and STRAUS, E. G., 1946, *Ann. of Mathem.*, **47**, 731.
 HÖNL, H., and PAPAPETROU, A., 1939 a, *Z. Phys.*, **112**, 512; 1939 b, *Ibid.*, **114**, 478; 1940, *Ibid.*, **116**, 153.
 MATHISSON, M., 1937, *Acta Physica Polonica*, **6**, 167.
 PAPAPETROU, A., 1939, *Praktika Acad. Athènes*, **14**, 540; 1940, *Ibid.*, **15**, 404.
 ROSENFELD, L., 1940, *Mem. Acad. Roy. Belgique*, **18**, no. 6.
 TETRODE, H., 1928, *Z. Phys.*, **49**, 858.
 TZU, H. Y., 1947, *Nature, Lond.*, **160**, 1947.

* It has been shown by Tzu (1947) that it is not possible to derive Blackett's formula for the magnetic field of the earth from a unified theory containing the constants G and c only. The necessity of introducing the new constant λ in equation (21) shows that our considerations do not contradict the conclusions of Tzu.

LXXXVIII. *The Determination of the Charge of Heavy Particles emitted during the Explosive Disintegration of Nuclei.*

By S. O. C. SÖRENSEN,

The H.H. Wills Physical Laboratory, University of Bristol *†.

[Received July 27, 1949]

[Plates XIV.–XVI.]

SUMMARY.

The charge of heavy nuclear fragments, ejected during the explosive disintegration of nuclei, can be precisely determined if the length of the track in a photographic emulsion is of the order of 1500μ or greater. The method is based on the determination of the number of δ -rays per unit length which contain more than a certain number of grains. The maximum observed energies of ejection of light nuclei, lithium, beryllium and boron are of the order of 500 MeV. The nuclear processes which lead to the emission of heavy fragments with such great energy remain to be explained.

INTRODUCTION.

In experiments with photographic plates exposed to cosmic radiation, several reports (Heitler, Powell and Fertel 1939, Hodgson and Perkins 1949, Bonetti and Dilworth 1949) have been given of the observation of heavy nuclear "fragments", ejected during the explosive disintegration of nuclei. In a recent extensive analysis of cosmic-ray stars produced in Kodak NT4 plates, exposed on the Jungfraujoch (3500 m.) and in high-altitude balloon flights, a few events of this type have been found in this laboratory. The most characteristic feature of the tracks is the large number of "knock-on" electrons or δ -rays, which the heavy particles produce in their passage through the emulsion. The study of the distribution of these δ -rays along the tracks provides a rather reliable method for determining the charge of the fragments. The new "electron-sensitive" plates, which give observable tracks of all electrons with an energy above >10 keV, are particularly suitable for such a study. In this paper, an account is given of the identification of lithium, beryllium and boron nuclei, with energies up to 500 MeV., emitted during the explosive disintegration of nuclei.

* On leave from the Physical Institute, University of Oslo.

† Communicated by Professor C. F. Powell, F.R.S.

THE DELTA-RAY METHOD.

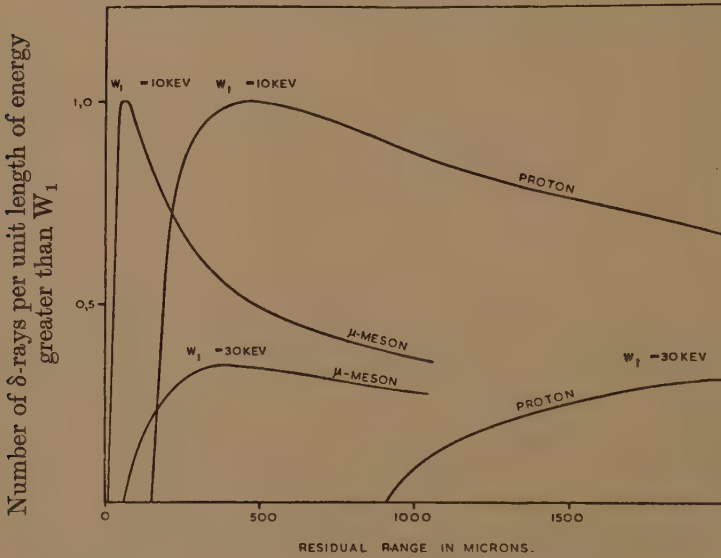
The method of identifying the charge of heavy nuclear particles, by studying the distribution of δ -rays along their tracks, has been developed mainly in order to study the heavy component of the primary cosmic radiation (Freier, Lofgren, Ney and Oppenheimer 1948, Bradt and Peters 1948). A similar method can be used in the determination of the charge of heavy "splinters" from cosmic-ray stars.

When a heavy particle, with charge Ze and velocity v , passes through the emulsion, the number, $d\nu$, of δ -rays with energies in the interval from W to $W+dW$, which it produces per unit length of its trajectory, is given by the formula

$$d\nu=2\pi Ne^4Z^2dW/mv^2W^2, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where m is the mass of the electron, and N the number of electrons per cm^3 of the emulsion. The total number of δ -rays per $\text{cm}.$, ν , which have

Fig. 1.



Theoretical δ -ray distributions for μ -mesons and protons.

The curves have been calculated from eq. (2) in the text, normalized to have a maximum value equal to unity for $W_1=10\text{KeV}$.

an energy greater than W_1 , is then given by integrating equation (1) between the limits from W_1 , to the maximum energy $W_2=2mv^2$, which a heavy particle with velocity, v , can transfer to an electron. The integration gives

$$\nu(W_1, v)=\frac{2\pi Ne^4Z^2}{mv^2}\left(\frac{1}{W_1}-\frac{1}{2mv^2}\right). \quad . \quad . \quad . \quad . \quad (2)$$

Equations (1) and (2) are only valid when the particle is moving at non-relativistic velocities. In what follows, $\nu(W_1, v)$ will be referred to as

the "density" of δ -rays along the trajectory. It is a quantity which clearly depends on the arbitrarily chosen value of W_1 , and which is subject to normal statistical fluctuations.

From (2), it follows that at points where particles with different values of the charge, Z , have the same velocity, $v(W_1, v)$ will vary as Z^2 ; and that the distributions of the values of $v(W_1, v)$ along the tracks of all particles will, apart from statistical fluctuations, be similar in form. Below a certain velocity, given by $v = \sqrt{(W_1/2m)}$, the particle will not produce any δ -rays with an energy greater than W_1 . Above this critical "cut-off" the density of δ -rays will increase at a rate which will depend on the variation of the velocity along the track. It will reach a maximum when the heavy fragment has a velocity $v = \sqrt{W_1/m}$, and then vary approximately as $1/v^2$ as the second term in the bracket of equation (2) becomes nearly constant. These relations are illustrated in figs. 1-3.

Fig. 2.

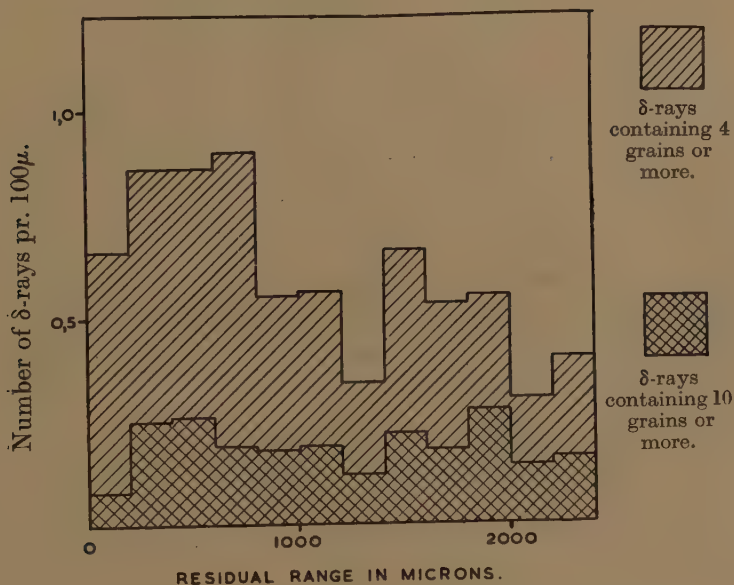
Experimental δ -ray distributions for μ -mesons.

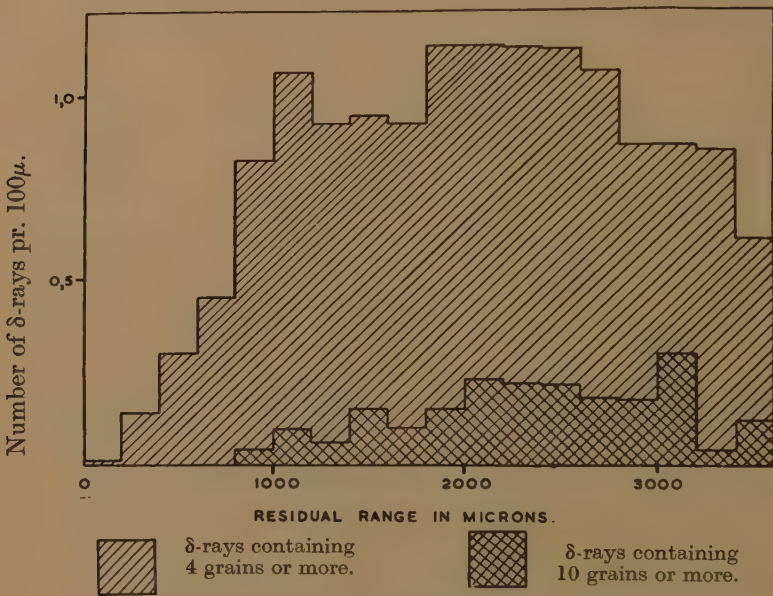
Fig. 1 gives the theoretical δ -ray distributions for μ -mesons and protons, calculated from formula (2) and the experimental range-velocity relations for these particles.

Fig. 2 gives the mean values of the observed frequency of occurrence of δ -rays along the tracks of 30 μ -mesons, all of which show the tracks of decay-electrons. For each of these particles, counts were made in successive elements of length 100μ , starting from the end of the range. No reliable measurements have yet been published of the energy of electrons of very short range in the NT4 emulsions, and the δ -rays have therefore been classified according to the number of grains which they contain.

For the mesons, the "cut-off" distance is negligible, for a value of W_1 corresponding to δ -rays of less than 10 grains. Further, the residual range of the particle, when the corresponding value of ν is at a maximum, is short, in accordance with the fact that the velocity of the particle is still relatively high near the end of its range.

Fig. 3 gives the result of similar measurements on the tracks of 20 protons. The "cut-off" effect, and the displacement of the maximum towards greater values of the residual range are immediately apparent. Further, it will be seen that the maximum values of ν , for mesons and protons, are approximately equal; a result which is in accordance with the equality of the charge of the two types of particles.

Fig. 3.



Experimental δ -ray distributions for protons.

These observations can now be employed for determining the charge, Z_2 , of a particle, in the following way: Counts are made, in equal intervals of length along the track, of all δ -rays which contain more than a given number of grains. The resulting distribution will, after a certain "cut-off", increase to a maximum and, then, slowly decrease. The maximum value of ν for this particular particle, ν_2 , may then be compared with that obtained from similar observations on the tracks of particles with a known charge Z_1 , such as those for mesons or protons given in fig. 2 and fig. 3. The charge is then simply given by the condition that

$$\frac{\nu_2(W_1)}{\nu_1(W_1)} = \frac{(Z_2^2)}{(Z_1^2)} \quad \dots \dots \dots (3)$$

For the comparison of the δ -ray distribution of an unknown heavy fragment with that of a particle with a known charge, it is an advantage to use α -particles, lithium or other light nuclei for calibration. Very few suitable tracks of such particles have been found, however, and we have been forced to use the δ -ray distribution in tracks of protons and mesons. In order to be sure of counting the δ -rays along the tracks of particles of the same type, only mesons which stopped in the emulsion and of which the tracks of decay-electrons could be distinguished, were chosen for measurement. It is certain that at least a very large fraction of such particles are μ -mesons of mass $216 \pm 3m_e$.

The main difficulty arising from the use of protons and mesons for calibration is that low-energy δ -rays, which can easily be distinguished when they originate from points in the thin "core" of the track of these lighter particles, have a tendency to be "lost" when associated with the thicker tracks of heavy fragments. Such an effect certainly exists but is found to be not very serious in dealing with particles with $Z < 6$ if δ -rays of sufficient length are chosen for counting.

IDENTIFICATION OF HEAVY SPLINTERS.

The material which has been analysed is made up of one track of a boron nucleus; one of a beryllium nucleus; five of lithium nuclei; and those of a number of α -particles. The boron fragment was emitted from a star of 9 branches and had a length of 1100μ , corresponding to an energy of about 250 MeV. The beryllium fragment was emitted from a star of 14 branches and had a length of 750μ , corresponding to an energy of about 140 MeV. The lithium fragments were emitted from stars which varied considerably in size, and the tracks had lengths up to $20,000\mu$, corresponding to maximum energies of about 500 MeV. A collection of micro-photographs of the different types of particles is shown in Plate XIV., whilst Plate XV. shows the track of the beryllium nucleus photographed with a greater magnification. The increase in the intensity of δ -rays with increasing charge is immediately apparent.

The identification of the nuclei depends on the measurements represented in fig. 4. The total number of δ -rays per 100μ , each of which contains 4 grains or more, is plotted in intervals of 200μ from the end of each track. This particular value for the minimum number of grains in the tracks of δ -rays, which are counted, has been found, by experience, to give the most consistent results. It corresponds to an electron of about 16 keV. If the limit is lowered to 3 grains or less, it becomes increasingly difficult to count all the low-energy δ -rays along the thick "core" of the tracks of the boron and beryllium fragments. On the other hand, an increase of the lower limit results in a large decrease in the total number of δ -rays counted; and this involves larger statistical fluctuations in the distribution curves showing the variation of ν along the track.

For the identification of the charge, we use the relations mentioned above. The maximum value of ν for the particles with charge $Z=1$

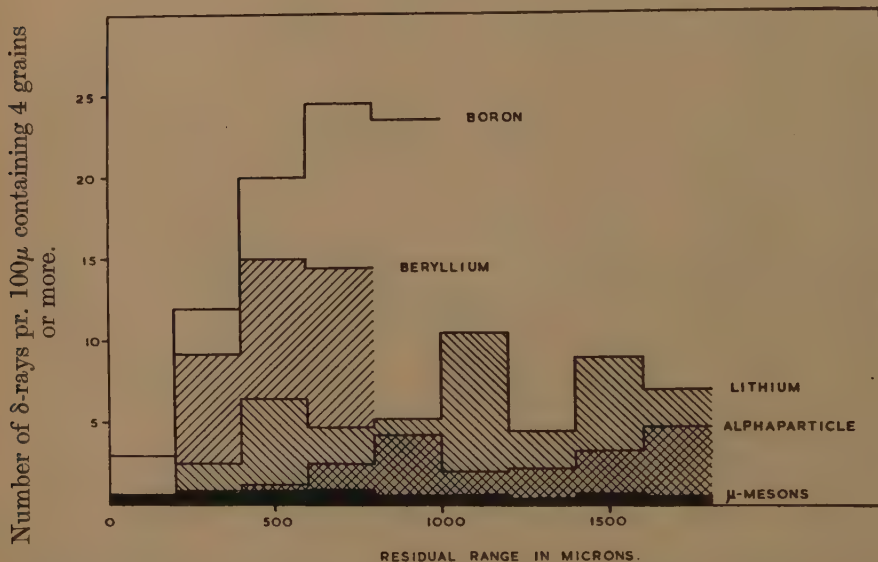
(mesons and protons) are found (figs. 2-3) to be : 1 ± 0.2 δ -rays, per 100μ , containing 4 grains or more ; *viz.* $\nu(4) = 1.0 \pm 0.2$. The corresponding expected maximum values, ν_{\max} for the other particles are shown in Table I.

TABLE I.

	Z	$\nu_{\max} = (1 \pm 0.2)Z^2$
α -particle	2	4 ± 0.8
lithium	3	9 ± 1.8
beryllium	4	16 ± 3.2
boron	5	25 ± 5.0

The maximum values of ν , in the observed distributions given in fig. 4, can be seen by inspection to be equal, within the limits given in Table I., to those corresponding to the assumed value of the charge. The results for the particular track which has been chosen as representative of a lithium nucleus shows exceptionally large statistical fluctuations. Even so, the values of ν appear to be too large to allow us to attribute the track

Fig. 4.



Experimental δ -ray distributions for nuclear fragments from cosmic ray stars. to an α -particle ; whilst to identify it as a beryllium nucleus is definitely impossible. The limited amount of material hitherto available suggests that a track should be at least 1500μ long if it is to provide sufficient evidence to decide between the possibilities that it was produced by an α -particle or by a lithium nucleus. In difficult cases, similar counts employing different values for the minimum number of grains in the δ -rays can be useful ; also the positions of the " cut-off " points along the tracks, for δ -rays of 6 grains or more, and 10 grains or more, etc., can give additional indications of the value of the charge.

In making the above analysis, the effect of capture and loss of charge, which begins to play a rôle towards the end of the range of a particle, has been ignored. This seems to be justified because, in the case of nuclei with $Z < 6$, such effects are confined to the last $\sim 25\mu$ of the range. The particle has then reached such a velocity that it can no longer produce δ -rays with an energy as great as the minimum value accepted for measurement. This condition is, however, not fulfilled in the case of nuclei of heavier elements. Photo-micrographs of the track of such a particle are reproduced in Plate XVI.

The track shown in Plate XVI. is an example of an heavy nucleus of the primary cosmic radiation, which was observed in an assembly of plates exposed at a maximum altitude of 130,000 feet. The particle entered the assembly in a direction inclined at an angle of about 30° to the vertical; and it was brought to rest in the emulsion of the 22nd plate which it traversed. By counting the δ -rays along its track, in regions in which the particle was moving at relativistic velocities, its charge was found to be $35 \pm 3e$. It will be seen that in the 450μ length of track observed in the last two contiguous emulsions through which it passed, plates 21 and 22, there is a continuous decrease in the specific ionization mostly due to the process of "capture and loss". In reaching these plates from plate 20, the particle had to pass through the glass of two plates, and it was in this unobserved region of its trajectory that the specific ionization of the particle was at its maximum value. Such heavy nuclei will already have started to capture electrons before the rate of production of δ -rays of energy 16 keV. has reached its maximum value. The "cut-off" effect and the effect of the electron capture, both of which result in a gradual decrease in track thickness towards the end, will, in these cases, partly "overlap", and the correct picture of the "thin down" phenomenon can hardly be given before a satisfactory theory of energy loss of multiply charged ions of low velocity exists. The method we have employed for the identification of lighter nuclei cannot, therefore, be applied in such cases without modifications.

The question whether the δ -ray method, besides giving the charge of a particle, is capable of allowing us to decide its mass, is also of interest. Different isotopes will have different values of the velocity for the same residual range. This results in a different distribution of the δ -rays, a displacement occurring in the position of the "cut-off" points, and of the maximum value of ν , etc. (The magnitudes of the maxima are of course not influenced by the change in the mass as long as we consider particles with the same charge.) An analysis of the differences to be expected between the δ -ray distributions in the tracks of neighbouring isotopes indicated, however, that, owing to statistical fluctuations, no unambiguous assignment would be possible. Nevertheless, a comparison between the theoretically expected positions of the maxima in fig. 4, and those actually observed, showed that the masses of the ejected nuclei were approximately equal to those of the stable isotopes with the same

charge. In making these estimates, 4 grains were assumed to correspond approximately to 16.5 keV. That the ejected particles were most probably stable or long-lived isotopes, is confirmed by the fact that none of them showed the tracks of decay-electrons at the end of their range. This cannot be a general feature of the heavy nuclear splinters emitted from stars, for other experimenters have observed the track of a Li^8 nucleus of range 500μ (Dilworth and Bonetti 1949). A detailed study of the relative frequency with which different nuclear types are emitted will possibly have a bearing on the interpretation of the mechanism of their ejection.

A point worth mentioning concerns the positions of the maxima and "cut-off" points in the δ -ray distributions for the different types of particles. If, for example, we consider δ -rays with energy greater than 16.5 keV., the positions of these points are given in Table II.

TABLE II.

Particle	"Cut-Off" Points	Maxima
μ -meson	40μ	134μ
proton	349μ	1163μ
α -particle	349μ	1163μ
lithium (7)	272μ	905μ
beryllium (9)	197μ	655μ
boron (11)	154μ	512μ

(Protons and α -particles have δ -ray distributions of the same form, apart from small differences due to "capture and loss", according to the equality of the factor M/Z^2). It was mentioned above that the tracks of α -particles or lithium nuclei ought to be longer than 1500μ , in order to ensure correct identification of the particles. For heavier nuclei, the maxima will be more pronounced as a result of the more rapid variation of the velocity with range. As shown in Table II., the distances of the maxima and "cut-off" points from the end of the range decrease with particles of increasing mass, a fact which depends on the decrease in the residual range, with increasing charge, for particles with the same velocity. These relations make a determination of the charge of a heavy particle possible even if its range is less than 1000μ .

DISCUSSION.

The available material is, of course, too small to allow us to draw any general conclusions concerning the processes which lead to the emission of heavy splinters from cosmic ray stars. It may, however, be of interest to indicate some points which characterize the observed events.

If it is legitimate to apply the usual methods of calculating range-energy relations, we are dealing with particles with a kinetic energy of the order of 300 MeV. These energies can hardly be explained in terms of ordinary evaporation or fission processes. The most reasonable assumption is to



Examples of the tracks of light nuclei ($Z=2, 3, 4$ and 5) emitted during the explosive disintegration of a nucleus.

To face page 954.



Track of a Beryllium nucleus, emitted from a nuclear explosion. The track starts at the top left and ends bottom right.



Track of a Bromine nucleus ($Z=35\pm 3$) which passes through the emulsion of 21 plates and ends in that of the 22nd, at the point marked *e*. The region of maximum ionization occurred in the glass of plate 21. Some distortion of the emulsion has occurred.

attribute the heavy splinters to some sort of "knock-on" effect produced by fast nucleons. In most of the events, the ejected heavy particles have been found to be moving in a direction making an angle of less than 90° with the downward vertical. Further, in some cases (for example, the beryllium nucleus in Plate XV.) there is a relativistic particle, of charge e , probably an incident proton, which moves in a direction making only a small angle with that of the heavy particle. It is difficult, however, to understand the process in terms of a collision of a relativistic proton or neutron with an aggregate of nucleons inside a nucleus, which leads to the ejection of the aggregate rather than to its complete disintegration. Another point which indicates a more complicated process is the occurrence of the simultaneous ejection of several heavy fragments. One star has been observed, in which at least two lithium fragments are emitted with great energy, and both the beryllium and the boron nuclei, of which the tracks are reproduced in Plate XIV., are accompanied by a fast lithium or α -particle, which move in directions making angles of the order of 90° with that of the heavy particle.

An interpretation of these nuclear processes can therefore hardly be attempted before more material is available. We must also leave open the question of the relative frequency of occurrence of the different types of fragments. We only mention that the events described have been found in the examination of about 6000 "stars".

Experiments are now in progress to examine the possibility of determining the mass of nuclear fragments by a study of their small angle Coulomb scattering.

ACKNOWLEDGMENTS.

I should like to express my gratitude to Professor C. F. Powell, F.R.S., for suggesting the experiment, for offering the facilities to carry out this work, and for his continued interest and help; and I take this opportunity of thanking the group of research workers and team of microscope observers of this laboratory for many discussions and suggestions. I am indebted to the Royal Norwegian Industry Department and Norges Teknisk Naturvitenskapelige Forskningsraad for a scholarship for studying nuclear physics in England.

REFERENCES.

- BONETTI and DILWORTH, 1949, *Phil. Mag.*, [7], **40**, 585.
BRADT and PETERS, 1948, *Phys. Rev.*, **74**, 1828.
FREIER, LOFGREN, NEY, and OPPENHEIMER, 1948, *Phys. Rev.*, **74**, 1818.
HEITLER, POWELL, and FERTEL, 1939, *Nature, Lond.*, **144**, 283.
HODGSON and PERKINS, 1949, *Nature, Lond.*, **163**, 439.

LXXXIX. *The Energy Distribution of Slow Positive Ion Beams.*

By G. D. YARNOLD and H. C. BOLTON.

University of Nottingham *.

[Received July 11, 1949.]

SUMMARY.

Consideration is given to the general problem of the selection of an ionic beam of small solid angle by a small hole in an electrode bounding an ionized region. The distribution of energy in the beam is determined principally by the character of the electric field obtaining in the ionized region and by the variation of the total collision cross-section of the ions as a function of velocity. The cathode dark space of a cold-cathode discharge is characterized in general by an electric intensity which increases rapidly as the cathode is approached. Under these conditions the beam is relatively rich in high energy ions, and, provided the mean free path is comparable with the length of the dark space, the distribution curve has a pronounced peak near to the high-energy limit. Assuming the field strength in the dark space to obey Aston's law, and the ionic cross-section to be inversely proportional to the energy, distribution curves are computed which show a marked similarity to those observed experimentally.

INTRODUCTION.

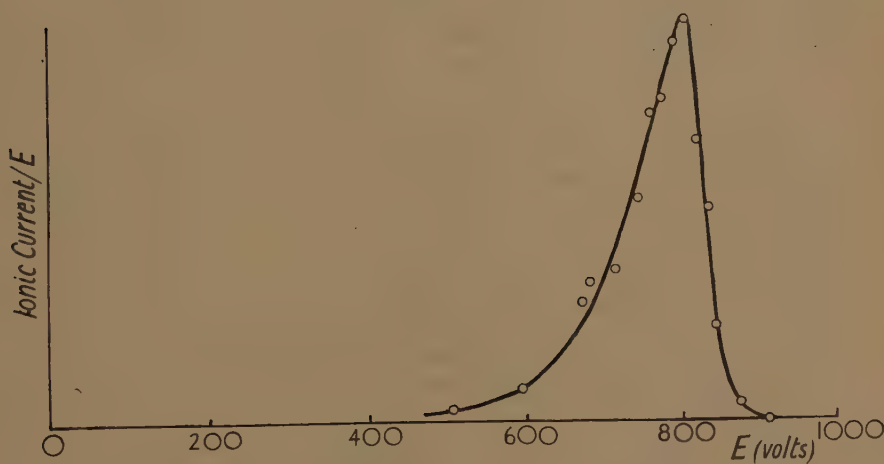
It is common knowledge that the energies of the canal rays from a discharge tube operating at low pressure cover a wide range and have an upper limit corresponding approximately to the difference of potential applied to the tube. The distribution of energies in the positive ion beam, which may be inferred both from the relative blackening of the different parts of the parabola of J. J. Thomson's early investigations and also from other experiments in which the beam is bent magnetically (*Handbuch* 1927), is characterized frequently by a slowly rising curve leading to a sharp peak which occurs at somewhat less than the maximum energy of the beam. In such investigations the pressure throughout the tube is normally of the order 10^{-3} mm., the applied potential difference of the order 10,000 volts and the mean free path comparable with the separation of the electrodes.

In the course of preliminary tests of a new kind of electrostatic separator (Yarnold and Bolton 1949) we have observed a similar distribution of energy in a beam of positive ions issuing from a small hole in the cathode

* Communicated by Professor L. F. Bates.

of a low-voltage discharge in hydrogen at a pressure of 1 mm. The diameter of the discharge tube was 2.3 cm. and the electrodes were 12 cm. apart. The iron cathode was pierced centrally by a hole, approximately 0.01 cm. in diameter, leading to the vessel containing the separator in which a good vacuum was maintained. The constant pressure was maintained in the discharge tube by means of a suitable leak, and the discharge current of the order 10 milliamperes was kept constant by a saturated diode. Under these conditions the difference of potential across the tube was less than 1000 volts. The separator, described previously, admitted only a narrow cone of positive ions from the hole in the cathode. The beam emerging from the separator was practically homogeneous in energy and was received in a Faraday cylinder; and by application of a series of suitable potentials in turn to the separator the

Fig. 1.

Energy distribution of positive ion beam. (Hydrogen, $p=1$ mm.)

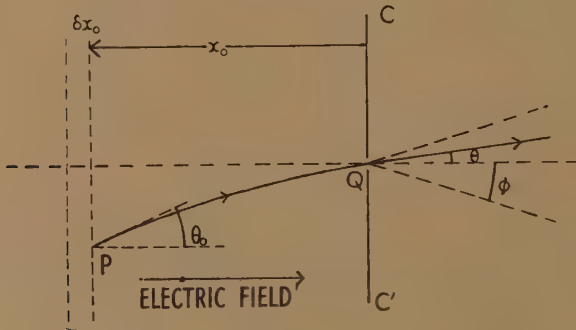
energy distribution of the original beam was studied. Since the energy range δE passed by the separator is proportional to the energy E , the true energy distribution of the beam is obtained by dividing the observed ionic currents by E . (Kollath 1936). A typical distribution is shown in fig. 1, where the same general characteristics appear as in the case of canal rays of comparatively high energy. Although the pressure of the gas in the discharge tube is relatively high in our experiments, yet the beam consists principally of ions having energies only slightly less than the difference of potential across the discharge. At this pressure the beam of high energy ions represents about one five-hundredth of the total current falling at all angles on the area of the hole. In agreement with other observers (Aston 1919, Brewer and Miller 1932) it was found that the high energy group of ions becomes a relatively larger fraction of the total current as the pressure in the discharge tube is reduced.

The theory of the selection of a positive ion beam by a hole in the cathode of a discharge tube has received little attention. While it is recognized that the fastest ions are those which have travelled without collision across the whole cathode dark space (Thomson 1924), no attempt has been made to account for the general form of the energy distribution. Accordingly, the purpose of this paper is to discuss the general problem of the passage of a narrow cone of positive ions through a hole in a cathode from any region containing an ionized gas, and in particular to account for the slowly rising part of the experimental distribution curve.

SELECTION OF A BEAM BY A SMALL HOLE.

Let the plane CC' (fig. 2) represent an indefinitely thin cathode separating a region to the left in which the gas pressure is p from a region to the right which is completely evacuated, and let the cathode be pierced at Q by a small hole of area A . Assuming axial symmetry, we may suppose that, either by the aperture of the separator or in some other way, the beam is limited to those ions which pass through the hole at angles less than ϕ .

Fig. 2.



In practice ϕ seldom exceeds 0.01 radian. Consider now an ion which passes through the hole at an angle θ having previously made a collision at P distant x_0 from the cathode. If the energy of the ion immediately after collision at P is E_0 , its energy at Q and at all points to the right of CC' may be written

$$E = E_0 + V_0, \quad \dots \dots \dots (1)$$

where V_0 represents the potential at P relative to the cathode. Since we are concerned only with ions moving at small inclinations to the direction of the field, we may easily show that

$$\frac{E_0}{E} \simeq \frac{\cot^2 \theta}{\cot^2 \theta_0} \simeq \frac{\theta_0^2}{\theta^2}, \quad \dots \dots \dots (2)$$

where θ_0 is the initial inclination of the path PQ .

The distribution in energy of the ions passing through the hole depends clearly on (1) the distribution in angle and energy of the ions immediately following their last collision, (2) the character of the electric field through which the ions pass after their last collision, and (3) the total collision cross-section of the ions, which is itself a function of velocity.

For sufficiently small inclinations to the direction of the electric field the distribution of the directions of motion of ions immediately after collision may be taken as isotropic. Thus, if N_0 is the total number of collisions taking place per second per unit volume in the space between the planes x_0 and $x_0 + \delta x_0$, the number of new free paths characterized by the energy range E_0 to $E_0 + \delta E_0$ and by the angular range θ_0 to $\theta_0 + \delta \theta_0$ may be written

$$\delta N_0 = k N_0 f_1(x_0, E_0) \delta E_0 \theta_0 \delta \theta_0, \quad (3)$$

where k is a constant. In general, the probability that an ion starting from P reaches the hole without further collision may be expressed as a function of the distance x_0 (approximately the path length) and the initial and final energies E_0 and E as

$$P_0 = f_2(x_0, E_0, E). \quad (4)$$

Thus, the number of ions passing through the hole per second having energies between E and $E + \delta E$ and inclinations between θ and $\theta + \delta \theta$ is

$$\delta N_\theta = \Sigma A P_0 \delta N_0 \delta x_0,$$

where $\delta E = \delta E_0 + \delta V_0 = \delta E_0 + X \delta x_0$, the summation being extended over all layers δx_0 for which equation (1) can be satisfied for the given value of E and a suitable value of E_0 . Hence

$$\delta N_\theta = \Sigma A k N_0 f_1(x_0, E_0) \delta E_0 f_2(x_0, E_0, E) \theta_0 \delta \theta_0 \delta x_0$$

which by (2) gives,

$$\delta N_\theta = \Sigma \frac{A}{X} \left(\frac{E}{E_0} \right) k N_0 f_1(x_0, E_0) \delta E_0 f_2(x_0, E_0, E) \theta \delta \theta \delta V_0$$

Finally, for the total number of ions passing through the hole per second, having energies in the range E to $E + \delta E$ and comprising the cone of angle ϕ , we obtain

$$\delta N = \Sigma \frac{A \phi^2}{2X} \left(\frac{E}{E_0} \right) k N_0 f_1(x_0, E_0) \delta E_0 f_2(x_0, E_0, E) \delta V_0. \quad (5)$$

If the functions f_1 and f_2 were known explicitly, together with the variation of V_0 and N_0 as functions of x_0 , then the number of ions passing through the hole in the angular range 0 to ϕ and energy range 0 to E is

$$N = \frac{A \phi^2 k E}{2} \int_{E_0=0}^{E_0=E-V_0} \int_{V_0=0}^{V_0=E} \frac{N_0}{X E_0} f_1(x_0, E_0) f_2(x_0, E_0, E) dE_0 dV_0. \quad . . . (6)$$

We have found it more convenient, however, to compute the energy distribution of the beam from equation (5) in several special cases where the functions occurring in the equation are known.

CASE (1). UNIFORM ELECTRIC FIELD AND CONSTANT COLLISION
CROSS-SECTION.

Here N_0 and f_1 are independent of x_0 provided the uniform field X extends sufficiently far from the cathode for the ions to have attained a steady state of motion before coming within a distance of one free path from the hole.

The energies of positive ions obeying classical laws of collision in a uniform electric field have been studied in a previous paper (Yarnold 1948), though all the necessary data for the evaluation of $f_1(E_0)$ were not published. Analysis of 964 collisions made by a group of ions which had already attained a steady state of motion in a uniform field X , however, yields the results exhibited in Table I., which shows the distribution in 5° ranges of the directions of motion immediately after impact, and the mean energy of each 5° group. The unit of energy adopted is the energy Xl acquired by an ion in moving through a distance equal to the mean free path l in the direction of the field. The angular distribution about

TABLE I.

Angular range	Number in range	Mean energy	Angular range	Number in range	Mean energy
0— 5°	9	1.15	45— 50°	64	0.87
5— 10°	9	2.01	50— 55°	83	0.80
10— 15°	22	1.46	55— 60°	79	0.65
15— 20°	37	0.95	60— 65°	57	0.50
20— 25°	39	1.18	65— 70°	69	0.52
25— 30°	47	1.44	70— 75°	55	0.43
30— 35°	54	1.22	75— 80°	50	0.46
35— 40°	63	0.93	80— 85°	39	0.37
40— 45°	71	1.05	85— 90°	37	0.28
			over 90°	80	

the direction of the electric field is approximately isotropic up to an inclination θ_0 of 30° . The mean energy of a group is subject to rather wide random fluctuations for inclinations less than 30° , where the number of collisions in any range is not large. Thereafter, the mean energy decreases fairly regularly with increasing angle. Nevertheless, the whole angular range 0 – 30° , comprising 163 collisions, may be used to make an approximate calculation of the energy distribution of ions of low inclination immediately after impact. The energies of this group are analysed in Table II.

Having regard to the relatively small number of collisions resulting in small inclinations θ_0 , the evidence for a peak in the energy distribution curve is not conclusive. For convenience the distribution is represented by writing

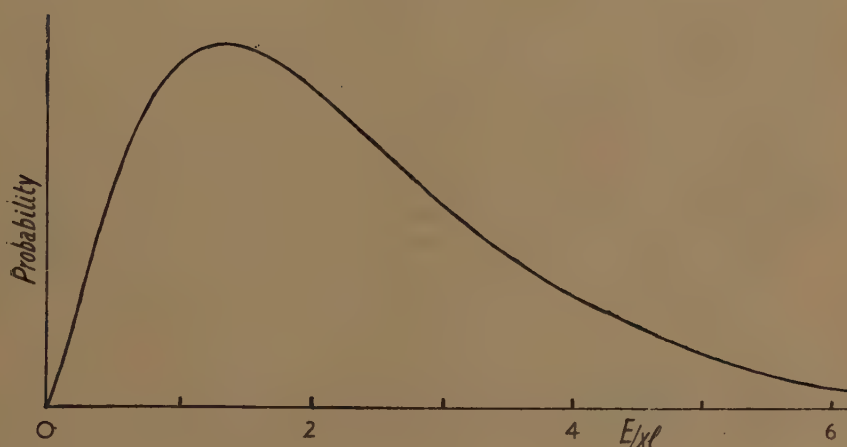
$$f_1(E_0) = \left(\frac{3}{4Xl} \right) \exp \left[-\frac{3E_0}{4Xl} \right] = \frac{1}{E_0} \cdot \exp \left[-\frac{E_0}{E_0} \right],$$

where \bar{E}_0 denotes the mean energy of an ion of small inclination immediately after collision. Moreover, for a constant collision cross-section $f_2(x_0, E_0, E)$ reduces to $\exp[-x_0/l] = \exp[-V_0/Xl]$. The energy distribution of the ions passing through the hole in the cathode (fig. 2) has been computed from equation (5) and is shown in fig. 3, where the

TABLE II.

Energy range	Number in range	Energy range	Number in range
0.0-0.2	13	1.8-2.0	3
0.2-0.4	19	2.0-2.2	5
0.4-0.6	20	2.2-2.4	4
0.6-0.8	20	2.4-2.6	11
0.8-1.0	15	2.6-2.8	6
1.0-1.2	7	2.8-3.0	5
1.2-1.4	11	3.0-4.0	5
1.4-1.6	11	4.0-5.0	2
1.6-1.8	4	5.0-6.0	2

Fig. 3.



Distribution for case (1).

ordinates are expressed in arbitrary units. The distribution approximates very closely to the empirical relation,

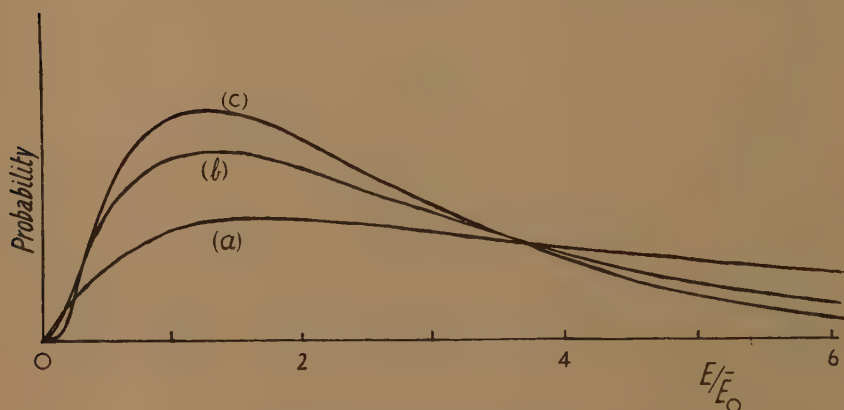
$$\delta N = C \left(\frac{E}{Xl} \right)^{4/3} \exp \left[-\frac{E}{Xl} \right] \frac{\delta E}{Xl} \dots \dots (7).$$

CASE (2). UNIFORM ELECTRIC FIELD AND VARIABLE COLLISION CROSS-SECTION.

Although many of the earlier investigations of collision cross-section concerned the passage of protons or alkali ions through the more common gases (Dempster 1927, Ramsauer and Beeck 1928), of far greater

importance is the passage of a positive ion through its own gas. The fundamental problem is the absorption of hydrogen ions in hydrogen, investigated first by Holzer (1930), by Ramsauer, Kollath and Lilienthal (1931), and most recently by Russell, Fontana and Simons (1941). Russell, Fontana and Simons have shown that the total cross-section α of protons in hydrogen at 1 mm. pressure is a function of the energy E of the ion, such that the product αE is constant throughout the range 0 to 200 electron volts. This appears to be a particular case of a more general law $\alpha E^m = \text{constant}$, where m takes different values according to the nature of the ion and the gas (Simons and Fryburg 1945, Simons and Unger 1945). These results, moreover, are in general agreement with the calculations of Massey and Smith (1933), who showed that at low energy only elastic scattering is important, while at higher energy absorption due to excitation and ionization occurs also.

Fig. 4.



Distributions for case (2).

(a) $Kp/X=2$. (b) $Kp/X=4$. (c) $Kp/X=6$.

Expressing the variation of the total cross-section of an ion as $\alpha E=K$, the probability of the occurrence of a collision in traversing a distance δx in a gas at pressure p is given by $\alpha p \delta x = Kp \delta x/E = (Kp/X) \delta E/E$, provided the ion is travelling approximately in the direction of the uniform field X . It follows that the probability that an ion shall attain an energy E from an initial energy E_0 without suffering collision is

$$f_2(x_0, E_0, E) = (E_0/E)^{Kp/X}.$$

The effect of a variable cross-section on the energy distribution of positive ions in the steady state of motion has not been investigated. Clearly, however, for ions obeying the law $\alpha E=K$, the general effect is to increase relatively the probability of the higher energy ranges. The distribution of energies immediately after collision, though modified, would be expected to differ only slightly from the distribution discussed in Case (1). We may assume for simplicity that the effect of the variable

cross-section is to raise the mean energy \bar{E}_0 of ions, whose directions of motion immediately after collision are inclined at small angles to the direction of the electric field, while leaving the form of the distribution of the energies about the mean energy substantially unchanged. Then $f_1(E_0) = (1/E_0) \exp[-E_0/\bar{E}_0]$. Thus, it is possible to compute the distribution of the energies of the ions passing through a hole in the cathode in terms of the unknown initial mean energy \bar{E}_0 of ions of small inclination, assuming numerical values for K and for the ratio X/p . Russell, Fontana and Simons give $K=850$ for protons moving in hydrogen with energies up to 200 electron volts. Fig. 4 shows the distribution as calculated from equation (5) for three integral values of the parameter Kp/X , which correspond to X/p equal to 425, 212 and 142 respectively, assuming $K=850$ for protons in hydrogen.

CASE (3). ASTON FIELD AND CONSTANT COLLISION CROSS-SECTION.

The conditions obtaining in the cathode dark space of an electric discharge between plane electrodes in a variety of gases were investigated by Aston (1911), who observed the deflection of an independently generated stream of cathode rays; and the electric field strength was found to be proportional to the distance from the sharp edge of the negative glow. Geddes (1926), employing the same method, confirmed Aston's work, but showed that the simple law for the variation of the field strength breaks down when the total cathode fall of potential exceeds about 1500 volts. Measurements with probes, though not entirely free from objection, have lent general support to these conclusions (Brown and Thomson, 1929).

Consider then a discharge in which the cathode fall of potential is V_1 and the width of the dark space is D . If Aston's law is assumed to hold, the potential V and the field strength X at distance x from the cathode are given respectively by

$$V = \frac{V_1 x}{D} \left(2 - \frac{x}{D} \right) \quad \dots \dots \dots (8)$$

$$\text{and} \quad X = \frac{2V_1}{D} \left(1 - \frac{x}{D} \right) = \frac{2V_1}{D} \left[1 - \left(\frac{E - E_0}{V_1} \right)^{\frac{1}{2}} \right] \quad \dots \dots \dots (9)$$

$$\text{since} \quad \frac{E - E_0}{V_1} = \frac{x}{D} \left(2 - \frac{x}{D} \right) = 1 - \left(1 - \frac{x}{D} \right)^2,$$

Moreover, the probability $\exp[-x_0/l]$ that an ion travels the distance x_0 to the cathode without collision, becomes

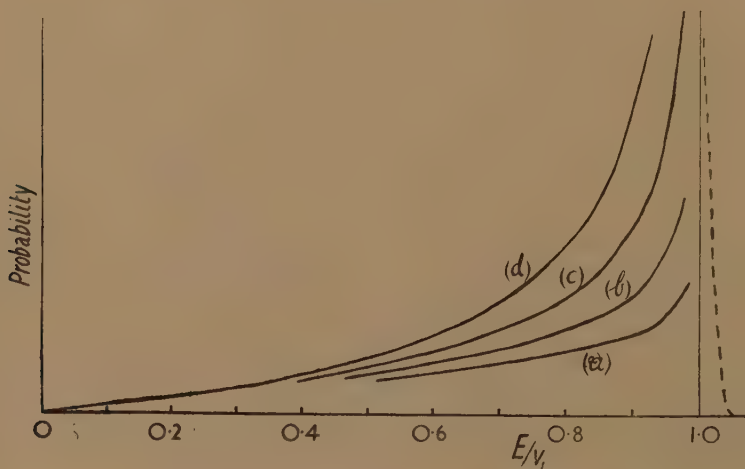
$$f_2(x_0, E_0, E) = \exp \left[-\frac{D}{l} \left\{ 1 - \left[1 - \left(\frac{E - E_0}{V_1} \right)^{\frac{1}{2}} \right] \right\} \right],$$

If we assume the cross-section for collision to be independent of the energy of an ion, the number of collisions N_0 taking place per second per unit volume is independent of x_0 . No data exist for an accurate evaluation of $f_1(x_0, E_0)$ in an electric field obeying Aston's law. However, neglecting the energy with which an ion enters the dark space from the negative

glow, the energy of an ion cannot subsequently exceed the difference in potential between its instantaneous position and the edge of the negative glow. Thus, the energy of an ion at distance x_0 from the cathode lies within the limits 0 and $(V_1 - V_0)$. In the absence of any known distribution law, we have assumed that all energies in this range are equally probable. Thus, $f_1(x_0, E_0) = 1/(V_1 - V_0)$.

The energy distribution of the ions passing through the hole in the cathode has again been computed from equation (5) on the basis of these assumptions, and is shown in fig. 5 for four different values of the ratio D/l , including the limiting case when $D/l \rightarrow 0$. The energies of the ions are expressed in terms of the ratio E/V_1 . Clearly, the distribution has an

Fig. 5.



Distributions for case (3). (Areas under curves not adjusted to equality.)

(a) $D/l=5$. (b) $D/l=2$. (c) $D/l=1$. (d) $D/l \rightarrow 0$.

infinity at $E=V_1$ since the electric field X has been assumed to vanish at the edge of the negative glow. In practice, the peak would be less sharp on account of the finite energy with which ions enter the dark space and the finite but small field obtaining in the negative glow. Thus, the maximum energy at the cathode is slightly in excess of V_1 and the cut-off is somewhat as indicated by the dotted line in fig. 5.

CASE (4). ASTON FIELD AND VARIABLE COLLISION CROSS-SECTION.

Expressing the variation of the total cross-section of an ion as in case (2), it is necessary to derive an expression for the probability $f_2(x_0, E_0, E)$ that an ion starting from a position x_0 with initial energy E_0 shall traverse the Aston field to the cathode without further collision, which it reaches with energy E . We have

$$\log f_2(x_0, E_0, E) = Kp \int_{x_0}^0 \frac{dx}{E_0 + V_0 - V},$$

where V_0 and V denote the potentials at x_0 and x respectively and

$E=E_0+V_0$. As in case (3), the initial energy E_0 cannot exceed V_1-V_0 , and all values of E_0 up to this limit must be assumed to be equally probable. Thus, writing $E_0=c(V_1-V_0)$, where $0 < c < 1$, we obtain

$$\log f_2(x_0, E_0, E) = Kp \int_{x_0}^0 \frac{dx}{B-V},$$

where $B=cV_1+(1-c)V_0$ and V is given by equation (8). Thus

$$\log f_2(x_0, E_0, E) = \frac{Kp}{V_1} \int_{x_0}^0 \frac{dx}{\frac{B}{V_1} - \frac{2x}{D} + \frac{x^2}{D^2}}$$

and

$$f_2(x_0, E_0, E) = \left[\frac{-\frac{B}{V_1} + \frac{x_0}{D} \left(1 + \sqrt{1 - \frac{B}{V_1}} \right)}{-\frac{B}{V_1} + \frac{x_0}{D} \left(1 - \sqrt{1 - \frac{B}{V_1}} \right)} \right]^{DKp/2V_1 \sqrt{1 - \frac{B}{V_1}}}$$

Substituting, $x_0/D = 1 - \sqrt{(V_1-V_0)/V_1}$, this may be expressed in a form convenient for computation as,

$$f_2(x_0, E_0, E) = \left[\frac{1 - (1-c) \sqrt{\left(\frac{V_1-V_0}{V_1} \right)} - \sqrt{(1-c)} \left\{ 1 - \sqrt{\left(\frac{V_1-V_0}{V_1} \right)} \right\}}{1 - (1-c) \sqrt{\left(\frac{V_1-V_0}{V_0} \right)} + \sqrt{(1-c)} \left\{ 1 - \sqrt{\left(\frac{V_1-V_0}{V_1} \right)} \right\}} \right]^{DKp/2V_1 \sqrt{(1-c) \left(\frac{V_1-V_0}{V_1} \right)}}$$

The distribution of initial energies is assumed to be given by

$$f_1(x_0, E_0) = 1/(V_1-V_0),$$

and the electric field X by equation (9) as in the computation of Case (3). Since the total collision cross-section is now variable, N_0 is a function of position in the dark space. Assuming that the mean energy of the ions crossing any given plane parallel to the cathode is proportional to the difference in potential between this plane and the edge of the negative glow, N_0 is proportional to $1/(V_1-V_0)$. The energy distribution of the ions passing through a hole in the cathode as computed from equation (5) is shown in fig. 6 for two values of the ratio $DKp/2V_1$, the energy again being expressed in terms of the ratio E/V_1 . The same considerations apply to the sharpness of the peak of the distribution curve as in Case (3).

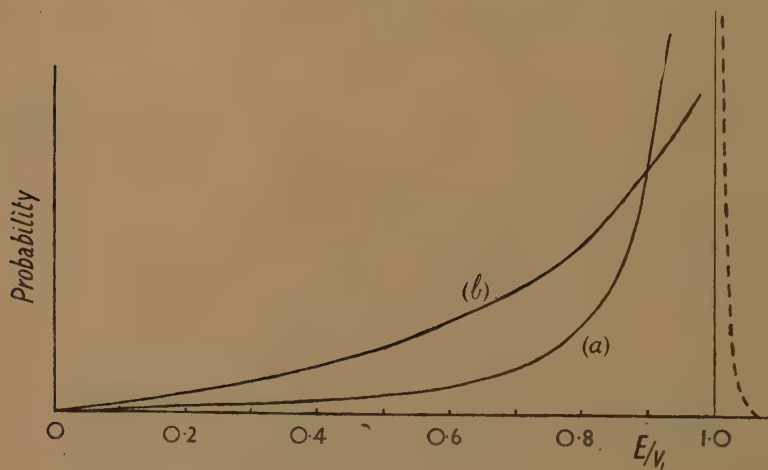
DISCUSSION.

Cases (1) and (2), in which the electric field is assumed to be uniform, clearly do not represent the conditions obtaining in an actual discharge tube, but are of interest for the sake of completeness. Whether the cross-section for collision is constant, or a function of velocity, the

maximum of the distribution curve occurs at the low energy end, whereas for a positive ion beam originating in a discharge tube the maximum occurs near the upper energy limit. When the total cross-section varies in accordance with the law $\alpha E = K$, the distribution for low values of the ratio X/p is not materially different from that obtained with a constant cross-section. At higher values of X/p , the distribution curve becomes flatter and the maximum moves towards the higher energies. The displacement of the maximum is in fact more pronounced than fig. 4 suggests, since \bar{E}_0 itself increases with increasing X/p .

Cases (3) and (4), in which an Aston field is assumed, give curves bearing a general resemblance to the energy distribution curve of a positive ion beam from a discharge tube, in that they show a slow rise to a sharp peak near the upper energy limit. In all cases it is clear that the height of the peak must be finite and the cut-off more gradual than in figs. 5 and 6,

Fig. 6.



Distributions for case (4).

(a) $DKp/2V_1 = 0.1$. (b) $DKp/2V_1 = 1$.

on account of the distribution of the initial energies with which the ions enter the dark space. Qualitatively, it can be seen from equation (5) that any distribution of the potential in the dark space which gives a relatively weak electric field near the negative glow together with a sufficiently strong field near the surface of the cathode (Geddes 1926), will give a rising energy distribution curve having a peak at the high energy end.

When the cross-section α is assumed to be constant (Case (3)), the sharpness of the peak in the energy distribution curve depends upon the ratio between the length of the dark space and the mean free path. When D is large in comparison with l , the probability of fast ions crossing nearly the whole length of the dark space without collision is small and the peak

of the curve is correspondingly reduced. Even when D/l tends to zero, the rise to the peak is much more gradual than that observed experimentally.

Case (4) most nearly represents the experimental facts when the total potential difference applied to the discharge is of the order of 1000 volts (fig. 1), provided that $DKp/2V_1$ is small. It should be observed that V_1/D represents the average field strength in the dark space, and hence that a low value of $DKp/2V_1$ implies a high effective value of X/p . It follows that the cross-section of an ion as it approaches the cathode decreases rapidly and the probability of an ion of the maximum energy reaching the cathode without collision is correspondingly increased. As compared with Case (3) the steepness of the rising curve is further enhanced by the variation in N_0 throughout the dark space, consequent on the variation in α . Although the work of Russell, Fontana and Simons was restricted to ions of less than 200 volts energy, the present extension of the range of application of the equation $\alpha E = K$ appears to be justified.

In Case (4) it is not possible to assign a definite mean free path to a group of accelerating ions independently of their initial energy E_0 but it is possible to estimate D . Assuming $DKp/2V_1 = 0.1$, and $K = 850$ for protons in hydrogen, we obtain $D/V_1 \simeq 2.4 \times 10^{-4}$ for $p = 1$ mm. of mercury. Comparison with fig. 6 suggests a value of $V_1 = 830$ volts for the experimental curve shown in fig. 1, giving $D \simeq 0.2$ cm., which is a reasonable value. Unfortunately, no corresponding experimental values of D , V_1 and the energy distribution exist for more exact comparison. The high energy tail of the curve in fig. 1 suggests that ions having energies up to 50 or perhaps 100 volts are entering the dark space from the negative glow. In other words, it appears that a small proportion of the ions travels the full distance of 10 cm. from the region of the anode to the negative glow without appreciable loss of energy.

An implicit cause of uncertainty in the interpretation of the experimental curves is the possibility of collision in the mouth of the hole, where the gas pressure drops rapidly but not, in practice, discontinuously. Moreover, the possibility of ionization within the dark space has been tacitly neglected. Any new ions so formed must have low initial energies and to a first approximation can be regarded as falling in the energy groups already considered. Nevertheless, the similarity of the rising curves of figs. 1 and 6 is sufficient justification for the general approach which we have adopted.

REFERENCES.

- ASTON, F. W., 1911, *Proc. Roy. Soc.*, **84**, 526.; 1919, *Ibid.*, **96**, 200.
 BREWER, A. K., and MILLER, R. R., 1932, *Phys. Rev.*, **42**, 786.
 BROWN, W. L., and THOMSON, E. E., 1929, *Phil. Mag.*, **8**, 918.
 DEMPSTER, A. J., 1927, *Phil. Mag.*, **3**, 115.
 GEDDES, A. E. M., 1926, *Proc. Roy. Soc. Edin.*, **46**, 136.
Handbuch der Physik, 1927, (Berlin: Springer), **24**, 78.
 HOLZER, R. E., 1930, *Phys. Rev.*, **36**, 1204.
 KOELLATH, R., 1939, *Ann. der Phys.*, **27**, 721.

- MASSEY, H. S. W., and SMITH, R. A., 1933, *Proc. Roy. Soc.*, **142**, 142.
 RAMSAUER, v C., and BEECK, O., 1928, *Ann. der Phys.*, **87**, 1.
 RAMSAUER, v C., KOLLATH, R., and LILLIENTHAL, D., 1931, *Ann. der Phys.*,
8, 702, 709.
 RUSSELL, A. S., FONTANA, C. M., and SIMONS, J. H., 1941, *J. Chem. Phys.*,
9, 381.
 SIMONS, J. H., and FRYBURG, G. C., 1945, *J. Chem. Phys.*, **13**, 216.
 SIMONS, J. H., and UNGER, L. G., 1945, *J. Chem. Phys.*, **13**, 221.
 THOMSON, G. P., 1924, *Proc. Roy. Soc. Edin.*, **44**, 129.
 YARNOLD, G. D., 1947, *Phil. Mag.*, **38**, 186.
 YARNOLD, G. D., and BOLTON, H. C., 1949, *J. Sc. Inst.*, **26**, 38.

XC. Notices of New Books and Periodicals received.

Surface Chemistry. (Various Authors.) (London: Butterworths, July 1949.)
 Price 25s.

THIS volume contains papers which were presented for discussion at a joint meeting of the Société de Chimie Physique and the Faraday Society in October 1947. They are grouped under the headings: Theoretical, Physical Chemistry (especially of monolayers at liquid surfaces), Films on Solid Surfaces, and Biophysical Chemistry, and give an indication of the progress made in recent years in the experimental techniques and theoretical understanding of adsorption phenomena.

The topical value of the book for specialists in its field is diminished, as some of the researches reported in it have been supplemented by work published since the meeting, and it seems a pity that there is no report of the discussion. However, it will no doubt be useful to workers interested in surface phenomena, as giving a broad survey of recent developments. G. W.

Introduction to Statistical Mechanics. By G. S. RUSHBROOKE. [Pp. 334.]
 (Oxford: University Press.) Price 21s.

THIS book is intended for third year Honours students of chemistry and graduates just starting research. The author examines in detail the three basic methods of statistical mechanics and their relations to one another. These methods can be summarized in the three formulæ:

$$\text{Entropy} = k \log \Omega,$$

where Ω is the number of complexions of an assembly with fixed energy;

$$\text{Available energy} = -kT \log (\text{partition function});$$

and

$$PV = kT \log (\text{grand partition function}).$$

The author uses mainly the first two methods, but he also points out the value of the grand partition function in avoiding the sometimes awkward problem of picking out the largest term in a series. These methods are used to calculate the specific heats and dielectric constants of gases and solids, and also to derive the law of mass action, adsorption isotherm, and some properties of imperfect gases and regular solutions.

This book provides a valuable introduction to statistical mechanics. When he has mastered it, the student should have a clear comprehension of the principles of the subject and be able to understand papers dealing with special aspects. Although primarily intended for chemists, it will also be of use to physicists, though for them it needs to be supplemented by a more detailed account of the Einstein-Bose and Fermi-Dirac statistics, which are only briefly dealt with here.

A. F. D.

From Euclid to Eddington. (The Tarner Lectures, 1947.) By SIR EDMUND WHITTAKER, F.R.S. [Pp. 212.] (Cambridge : University Press.) Price 15s.

THIS book contains a series of essays on the fundamental principles of physics. The concepts of space and time and their relation to the theory of relativity are discussed at some length. A section is devoted to the principles of quantum mechanics and another to Eddington's theory of the universe.

Any physicist will find it refreshing to read this book. For here we have the fundamentals of physics lucidly expounded, freed from the mass of detail which usually obscures them, and their mutual relations clearly shown.

A. F. D.

Formulæ and Theorems for the Special Functions of Mathematical Physics. By W. MAGNUS and F. OBERHETTINGER. Translated from the German by John Wermer [170 Pp.]. (Chelsea Publishing Co., N.Y., 1949.)

THIS translation from the German work of 1943 will be very welcome; containing as it does an invaluable collection of the formulæ of Bessel, Elliptic, Legendre and most other functions of mathematical physics.

NOTES FOR THE GUIDANCE OF AUTHORS SUBMITTING PAPERS FOR PUBLICATION.

Papers should be in typescript with double spacing. One side only of the paper should be used. MSS. should be as brief as is consistent with clarity. In particular the citation of elementary steps in a mathematical argument is to be avoided.

An abstract should always be provided. This should be as informative as possible, should be placed at the head of the paper and should not exceed 200 words in length.

USE OF CERTAIN MATHEMATICAL SYMBOLS.

In mathematical expressions appearing in the solid text the task of the compositor can be appreciably lightened in several ways, such as the use of the solidus and of the *exp* notation with a careful employment of the bracket. Thus, write $\sin (\theta/2)$ not $\sin \left(\frac{\theta}{2}\right)$; $(\sin \theta)/2$ not $\frac{\sin \theta}{2}$; $(a+b)/(c+d)$ not $\frac{a+b}{c+d}$; $a+b/c+d$ not $a+\frac{b}{c}+d$. Also $\sqrt{a^2+b^2}$ or $(a^2+b^2)^{\frac{1}{2}}$ not $\sqrt{a^2+b^2}$; $n!$ not \underline{n} ; and $3kT/2$ not $3/2kT$ —a common error. An example of the *exp* notation is Andrade's equation: $\eta=A[\exp (B/T)]$.

In a mathematical argument, however, the formulæ or equations should be written out on separate lines (*displayed* is the technical term) and numbered. The necessity for the use of the solidus is not so pressing.

ILLUSTRATIONS.

Line drawings should be made in Indian ink on Bristol board or tracing-cloth. Foolscap size should be the maximum, and unless the work is executed by a professional draughtsman the diagrams should be lettered lightly in pencil. Photographs are usually reproduced as plates, and should be not used unless absolutely necessary.

REFERENCES.

References should be made in the solid text by giving the author's name and year of publication in brackets. Thus: "It has been shown (Jones 1935) that" If Jones has published more than one paper in 1935, the papers should be referred to in order of date as (Jones 1935 a, b, . . .). References should be collected in alphabetical order of authors and placed at the end of the paper under the heading "References". Each reference should be of the form: Author's name; year of publication; abbreviated title of journal; series (if any) in square brackets; volume in Clarendon arabic type; page number. Thus: Jones, A. B., 1935, *Phil. Mag.* [7], **19**, 742.

ABBREVIATIONS.

The Royal Society recommends the following notation for multiples and submultiples of any unit:

10^6	10^3	1	10^{-1}	10^{-2}	10^{-3}	10^{-6}	10^{-9}	10^{-12}
M	k		d	c	m	μ	μm	$\mu\mu$

Further information may be obtained from:

- (1) The Royal Society's pamphlet "Notes on the preparation of papers communicated to the Royal Society";
- (2) Report of a Joint Committee on the Chemical, Physical and Faraday Societies on Symbols and Conventions,

XCI. *Plastic Distortion of Non-uniform Sheets.*

By R. HILL, M.A., Ph.D.

Cavendish Laboratory, Cambridge*.

[Received July 12, 1949.]

1. INTRODUCTION.

THE paper is concerned with the plastic straining of a flat sheet in its plane. The thickness of the sheet may be non-uniform in the stress-free condition or may become so as a result of plastic distortion. It will be assumed that the state of stress is plane, or nearly so. A state of stress is said to be plane with respect to Cartesian axes (x, y) when the stress components $\sigma_z, \tau_{xz}, \tau_{yz}$, in the z direction, are zero. If a plate of uniform thickness is loaded along its edge, by forces acting in its plane, the distribution of stress should be very nearly plane sufficiently far from the edge. If, during plastic deformation under these applied loads, the thickness does not remain uniform the state of stress may still be treated as approximately plane so long as dh/ds is small compared with unity, where h is the local thickness and s is the distance in any direction parallel to the surface.

2. THE STRESS EQUATIONS.

If $\sigma_x, \sigma_y, \tau_{xy}$ denote values of the stress components averaged through the thickness, the equations of equilibrium are (in the absence of body forces)

$$\frac{\partial}{\partial x}(h\sigma_x) + \frac{\partial}{\partial y}(h\tau_{xy}) = 0, \quad \frac{\partial}{\partial x}(h\tau_{xy}) + \frac{\partial}{\partial y}(h\sigma_y) = 0. \quad (1)$$

The averaged stresses may be inserted in the yield criterion, with only a small error. Let the criterion be

$$f(\sigma_x, \sigma_y, \tau_{xy}) = 0. \quad (2)$$

Suppose the stress components are given on some curve C in the (x, y) plane, and let the axes of x and y be taken respectively normal and tangential to C at some point P . We wish to find when the tangent at C is a characteristic direction for equations (1) and (2). Since the stresses are given along C , their y derivatives are known at P . The slope $(\partial h/\partial x, \partial h/\partial y)$ of the surface will normally be continuous, and so the equilibrium equations give $\partial\sigma_x/\partial x$ and $\partial\tau_{xy}/\partial x$. For the determination of $\partial\sigma_y/\partial x$ we have the equation

$$0 = \frac{\partial f}{\partial x} = \frac{\partial f}{\partial \sigma_x} \frac{\partial \sigma_x}{\partial x} + \frac{\partial f}{\partial \sigma_y} \frac{\partial \sigma_y}{\partial x} + \frac{\partial f}{\partial \tau_{xy}} \frac{\partial \tau_{xy}}{\partial x}.$$

* Communicated by the Author.

The elimination of p and τ_m between (4), (5) and (6) furnishes the (σ, τ) equation of the envelope. If the inclination of the tangent to E is denoted by ψ (regarded as positive when $|\tau|$ decreases with increasing tension σ), and the inclination of the tangent to F by χ , equation (6) expresses the fact that at corresponding points on E and F,

$$\sin \psi = \tan \chi. \quad (7)$$

Thus, only when $-\frac{1}{4}\pi \leq \chi \leq \frac{1}{4}\pi$ is there real contact between a stress circle for a plastic state and the envelope E; there are possible plastic states corresponding to circles lying entirely inside E.

It will now be shown that the stress equations are hyperbolic or elliptic according as the contact with the envelope is real or imaginary, respectively. The condition (3) for a characteristic is equivalent to

$$0 = \frac{\partial F}{\partial \sigma_y} = \frac{\partial F}{\partial p} \frac{\partial p}{\partial \sigma_y} + \frac{\partial F}{\partial \tau_m} \frac{\partial \tau_m}{\partial \sigma_y}.$$

$$\text{Now} \quad p = -\frac{1}{2}(\sigma_x + \sigma_y), \quad \tau_m^2 = \frac{1}{4}(\sigma_x - \sigma_y)^2 + \tau_{xy}^2;$$

$$\frac{\partial p}{\partial \sigma_y} = -\frac{1}{2}, \quad \frac{\partial \tau_m}{\partial \sigma_y} = \frac{-(\sigma_x - \sigma_y)}{4\tau_m} = \frac{-(\sigma_x + p)}{2\tau_m}.$$

Hence the condition for a characteristic is

$$\frac{\sigma_x + p}{\tau_m} = -\frac{\partial F}{\partial p} \bigg/ \frac{\partial F}{\partial \tau_m}. \quad (8)$$

Comparing (8) with (6) we observe that C is a characteristic when, at every point, the normal stress component acting across C corresponds to the point of contact with the envelope. Since there are two points of contact for a given plastic state, there are two characteristic directions; these are inclined at an angle

$$\frac{1}{4}\pi + \frac{1}{2}\psi$$

to the direction of the algebraically greater principal stress. Thus, the stress equations are hyperbolic if, and only if, the contact is real. When $\chi = \pm \frac{1}{4}\pi$ and $\psi = \pm \frac{1}{2}\pi$ the characteristics are coincident. These points on the locus F correspond to the maximum and minimum values of σ_1 and σ_2 ; the characteristics coincide with the axis of the numerically lesser principal stress.

It may be remarked in passing that this investigation of characteristics is applicable, without change, to plastic states in the plane (two-dimensional) deformation of a material whose yielding is influenced by hydrostatic pressure. The latter problem has been examined by Mandel (1942), who employed geometrical reasoning and also the standard method of the characteristic determinant.

We now specialize the yield criterion. For plane stress, von Mises' criterion reduces to

$$\sigma_1^2 - \sigma_1\sigma_2 + \sigma_2^2 = 3k^2, \quad (9)$$

where k is the yield stress in pure shear ($k=Y/\sqrt{3}$, where Y is the tensile yield stress). This equation, considered as a locus in the (σ_1, σ_2) plane, represents an ellipse (fig. 2). Referred to its principal axes the equation of the ellipse is

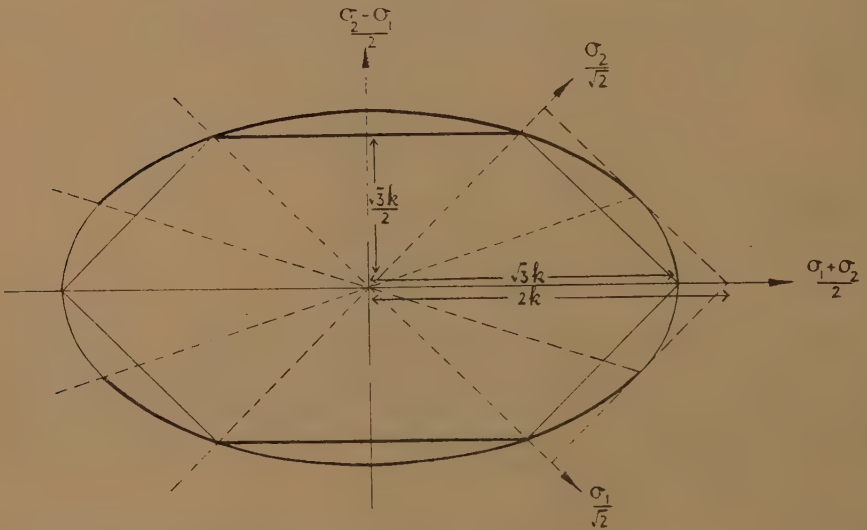
$$F(p, \tau_m) \equiv \frac{1}{3}p^2 + \tau_m^2 = k^2.$$

The envelope E may be shown to be

$$\frac{1}{4}\sigma^2 + \tau^2 = k^2.$$

The stress equations are hyperbolic for states corresponding to points on the thickened arcs of the ellipse in fig. 2. Over the remainder of the ellipse, where both $|\sigma_1|$ and $|\sigma_2|$ are greater than k (or, equivalently, where the maximum shear stress in the plane is numerically less than $\frac{1}{2}k$),

Fig. 2.



the equations are elliptic. The characteristic directions are coincident with the axis of the numerically lesser principal stress when the principal stresses have the values $\pm(k, 2k)$. The angle ψ may be shown, without difficulty, to satisfy the equation

$$\sin \psi = \frac{1}{3} \left(\frac{\sigma_1 + \sigma_2}{\sigma_1 - \sigma_2} \right); \quad \sigma_1 > \sigma_2. \quad . \quad . \quad . \quad . \quad . \quad (10)$$

These results were originally obtained by Sokolovsky (1945), though by a less elegant method.

Consider, next, Tresca's maximum shear stress criterion. The F locus is a hexagon inscribed in the Mises ellipse (fig. 2). When σ_1 and σ_2 have the same sign, the maximum shear stress lies outside the plane and the greater of $|\sigma_1|$ and $|\sigma_2|$ is equal to $2k$, where k is now written for $\frac{1}{2}Y$. When σ_1 and σ_2 have opposite signs, the maximum shear stress lies in the plane and $|\sigma_1 - \sigma_2| = 2k$. The envelope E coincides with F for stress

states such that $|\sigma_1 - \sigma_2| = 2k$, but for all other states it degenerates into the pair of points $(\pm 2k, 0)$. In the first case the equations are hyperbolic, with the maximum shear stress directions as characteristics. In the second case there is only a single characteristic, which coincides with the direction of the numerically lesser principal stress.

As we shall only be concerned with radially symmetric stress distributions here, it is unnecessary to obtain the differential relations holding along the characteristics. This will be left for a later paper dealing with more difficult problems.

3. THE VELOCITY EQUATIONS.

Let the (x, y) components of velocity averaged through the thickness of the sheet be denoted by (u, v) . To the usual order of approximation, these, together with the averaged stresses, may be inserted in the plastic stress-strain relations. It will be supposed, for simplicity, that Young's modulus has an infinitely great value, so that an element is rigid when stressed below the yield limit; furthermore, when an element is plastic the elastic component of the strain vanishes. The relations between stress and strain-rate may then be taken, with sufficient generality, as

$$\frac{\partial u}{\partial x} \bigg/ \frac{\partial g}{\partial \sigma_x} = \frac{\partial v}{\partial y} \bigg/ \frac{\partial g}{\partial \sigma_y} = \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \bigg/ \frac{\partial g}{\partial \tau_{xy}}, \quad \dots \quad (11)$$

where g is an invariant function of $\sigma_x, \sigma_y, \tau_{xy}$, and is known as the plastic potential.

Let C be a curve along which u, v , and the components of stress are given, and let the (x, y) axes be taken to coincide respectively with the normal and tangent to C at some point P . Then, if the velocity is continuous across C , $\partial u / \partial y$ and $\partial v / \partial y$ are known at P . Equations (11) suffice to determine $\partial u / \partial x$ and $\partial v / \partial x$ uniquely at P unless

$$\partial g / \partial \sigma_y = 0 \quad \dots \quad (12)$$

at P , so that $\partial v / \partial y = 0$. Thus C is a characteristic for the velocity components if it coincides at every point with a direction of zero rate of extension. There are two such directions through any point, but, because of the z component of strain, they are not generally orthogonal, or necessarily even real.

On comparing equations (3) and (12) it will be seen that the stress and velocity characteristics are identical if the plastic potential g is equal to the function f defining the yield criterion. This is so, in particular, when the Lévy-Mises relations

$$\frac{\partial u / \partial x}{2\sigma_x - \sigma_y} = \frac{\partial v / \partial y}{2\sigma_y - \sigma_x} = \frac{\partial u / \partial y + \partial v / \partial x}{6\tau_{xy}}, \quad \dots \quad (13)$$

are used together with von Mises' yield criterion (9). When the Lévy-Mises relations apply, the inclination $\frac{1}{4}\pi + \frac{1}{2}\psi$ of the velocity characteristics to the axis of the algebraically greater principal stress is given by (10).

If Tresca's yield criterion is used in conjunction with (13) the stress and velocity characteristics are in general distinct.

The amount of thickening or thinning of the sheet is determined from the condition for zero volume change. The rate of strain in the z direction, averaged through the thickness, is

$$\frac{1}{h} \frac{Dh}{Dt} = \frac{1}{h} \left(\frac{\partial h}{\partial t} + u \frac{\partial h}{\partial x} + v \frac{\partial h}{\partial y} \right),$$

where D/Dt is the operator denoting rate of change following an element. The equation of incompressibility is therefore

$$\left. \begin{aligned} \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{1}{h} \frac{Dh}{Dt} &= 0, \\ \frac{\partial h}{\partial t} + \frac{\partial}{\partial x}(hu) + \frac{\partial}{\partial y}(hv) &= 0. \end{aligned} \right\} \quad \dots \dots (14)$$

We may note in passing that, when (13) holds good, the sign of the rate of strain in the z direction is that of $-(\sigma_x + \sigma_y)$ or of p ; this follows from the fact that the ratios in (13) must be positive in order that the rate of plastic work is positive. Wherever the discontinuity in velocity gradient (*i. e.*, strain-rate) allowed by (11) or (13) occurs, there must also be one in $\partial h/\partial t$; this implies that as a result of the ensuing increment of strain the slope of the surface changes abruptly.

Now we are assuming that the part of the sheet which is not stressed to the yield limit is rigid. It therefore follows, by the properties of characteristics, that the plastic material is also rigid everywhere within the triangular area bounded by the plastic boundary and the intersecting velocity characteristics through its terminal points. A change in thickness is only possible outside such an area. This observation does not appear to have been made by any previous writer. If, on the other hand, the elastic component of strain were not disregarded, thickening strains could occur within the corresponding area, but would only be of an elastic order of magnitude.

4. EXPANSION OF A CIRCULAR HOLE IN A PLATE.

The only work on special problems of plane stress in which the progressive changes in thickness have been adequately incorporated in the solution, other than where the state of stress is uniform, appears to be that of Taylor (1948). Taylor examined the problem of the expansion of a hole by internal pressure in an infinite plate, starting from zero radius. In certain Russian work (Sokolovsky, 1946) on the drawing of thin strips through a die, under conditions of plane stress, the appreciable thickening that must occur is ignored and it is not shown that the proposed plastic region satisfies the velocity, as well as the stress, boundary conditions. Other investigations (Sokolovsky, 1946) on the plastic zones round holes of arbitrary shape in a tensioned plate are subject to the same limitations as similar work on plane strain, discussed in a paper by the author (1949); an additional defect is that the circumstances under which local thinning is initiated are not examined.

As an illustration of the foregoing results on characteristics we shall consider the enlarging of a circular hole from some finite radius in an infinite plate. Now the solution is really included (in a sense to be explained) in that for a hole expanded from zero radius, which is the problem investigated by Taylor. However, Taylor's solution is not entirely correct, and it is worth while to examine the problem afresh. A different method has been adopted, since it appears more advantageous to work with velocities rather than with displacements.

Consider a circular hole of radius a in a uniform plate, and let a gradually increasing pressure P be applied uniformly over the edge of the hole. While the plate is stressed below the yield limit the radial and circumferential components of stress are known from elastic theory to be

$$\sigma_r = -\frac{Pa^2}{r^2}, \quad \sigma_\theta = \frac{Pa^2}{r^2}.$$

Since the state of stress in every element is a pure shear, yielding begins first on the edge of the hole, at the pressure k . If the pressure is now raised further the material round the hole becomes plastic within some radius c . The stresses in the non-plastic material are obviously

$$\sigma_r = -\frac{kc^2}{r^2}, \quad \sigma_\theta = \frac{kc^2}{r^2} \quad (c \leq r).$$

We have to decide whether or not the plate immediately begins to thicken. As shown in the preceding section, the plate does not thicken in a certain area near the plastic boundary if the velocity equations are hyperbolic. Adopting the Lévy-Mises relations (13), this is so if the angle ψ , defined by (10), is real; this condition is certainly satisfied on the plastic boundary where $\sigma_1 + \sigma_2 = 0$ and $\psi = 0$. Hence the velocity characteristic at points on the plastic boundary are inclined to the radial direction at a finite angle, namely 45° . Thus the rigid part of the plastic region must extend over a finite annulus, its inner boundary being the circle where the velocity characteristics are coincident. In the rigid plastic annulus, then, the thickness is unaltered and the equation of equilibrium is simply

$$\frac{d\sigma_r}{dr} = \frac{\sigma_\theta - \sigma_r}{r}.$$

To facilitate the integration if the yield criterion of von Mises is adopted, we satisfy (9) by introducing a parameter θ , such that

$$\sigma_\theta = 2k \sin\left(\frac{1}{6}\pi + \theta\right), \quad \sigma_r = -2k \sin\left(\frac{1}{6}\pi - \theta\right), \quad \sigma_\theta - \sigma_r = 2k \cos \theta,$$

where θ is zero on the plastic boundary. Note that $\tan \theta = \sqrt{3} \sin \psi$. Substituting in the equilibrium equation:

$$\cos\left(\frac{1}{6}\pi - \theta\right) \frac{d\theta}{dr} = \frac{\cos \theta}{r},$$

where it has been supposed that k is a constant (no work-hardening).

Integrating:

$$\frac{c^2}{r^2} = e^{-\sqrt{3}\theta} \cos \theta.$$

The internal pressure producing a plastic region of radius c is given (Nadai, 1931) parametrically by

$$P = 2k \sin \left(\frac{1}{3}\pi - \theta_a \right), \quad \frac{c^2}{a^2} = e^{-\sqrt{3}\theta_a} \cos \theta_a.$$

θ_a becomes increasingly negative with increasing P , and the characteristics coincide when $\theta_a = -\frac{1}{3}\pi$, or $c/a = \rho$, where

$$\rho = \left(\frac{1}{2} \exp \frac{\pi}{\sqrt{3}} \right)^{\frac{1}{2}} = 1.751, \text{ approx.} \quad . \quad . \quad . \quad (15)$$

The corresponding value of the angle ψ is $-\frac{1}{2}\pi$, and the angle $\frac{1}{4}\pi + \frac{1}{2}\psi$ at which the characteristics are inclined to the direction of the algebraically greater principal stress (viz. σ_θ) is zero; thus the characteristics envelop the edge of the hole. The pressure required just to advance the plastic region to the radius ρa is equal to $2k$; the circumferential stress at the edge is then a compression of amount k . If a still greater load is applied, the plastic boundary continues to expand and the inner radius of the rigid annulus of plastic material is still the constant fraction $1/\rho$ (0.571) of the radius of the whole plastic region. The material inside the radius $0.571 c$ is not constrained to remain rigid by the material further out, and a thickening becomes possible. That such a thickening *does* occur is due to the fact that the plastic material cannot sustain a greater stress than $2k$, the pressure already applied; hence, if the load is increased, the plate must thicken to support it, and the pressure can be expected to fall.

Turning now to the corresponding analysis for Tresca's criterion, we have

$$\sigma_\theta - \sigma_r = 2k$$

in a certain annulus near the plastic boundary where the principal stresses have opposite signs. Thus, from the equation of equilibrium,

$$\sigma_r = -k \left(1 + 2 \ln \frac{c}{r} \right), \quad \sigma_\theta = k \left(1 - 2 \ln \frac{c}{r} \right).$$

When the internal pressure is raised to its maximum possible value $2k$ the plastic boundary is advanced to a radius ρa , where

$$\rho = \sqrt{e} = 1.649, \text{ approx.} \quad . \quad . \quad . \quad . \quad . \quad (16)$$

The application of a greater load causes a thickening, and the rigid annulus of plastic material is confined between the radii $0.607 c$ and c . That the velocity equations are hyperbolic in this annulus may be seen from (10), since the angle

$$\psi = -\sin^{-1} \left(\frac{2}{3} \ln \frac{c}{r} \right)$$

is real ($-\sin^{-1} \frac{1}{3} \leq \psi \leq 0$), when $\sqrt{ec} \leq r \leq c$. When a greater load is applied σ_θ becomes negative on the edge of the hole, and the yield criterion changes to

$$\sigma_r = -2k,$$

σ_θ varies discontinuously across the radius $0.607c$ from zero to $-k$, the value needed to preserve zero circumferential strain in the presence of a thickening. Just inside this radius, $\psi = -\frac{1}{2}\pi$ and the velocity characteristics coincide with the circumferential direction.

For the analysis of the ensuing distortion the equations of equilibrium, incompressibility, and Lévy-Mises are respectively

$$\left. \begin{aligned} \frac{\partial}{\partial r}(h\sigma_r) &= \frac{h(\sigma_\theta - \sigma_r)}{r}, \\ \frac{\partial h}{\partial c} + \frac{\partial}{\partial r}(hv) + \frac{hv}{r} &= 0, \\ \frac{\partial v}{\partial r} &= \left(\frac{2\sigma_r - \sigma_\theta}{2\sigma_\theta - \sigma_r} \right) \frac{v}{r}, \end{aligned} \right\} \dots \dots \dots (17)$$

where v denotes the radial velocity with the parameter c as the time-scale. This system of equations is hyperbolic, with $dc=0$ and $dr-vdc=0$ as characteristic directions in the (r, c) plane. If the system is considered by itself, and not as a specialization of the general (x, y) equations, the fundamental reason for the existence of a rigid annulus of plastic material is largely obscured. Similarly, the inner boundary of the rigid annulus no longer appears as the locus of points where the velocity characteristics coincide, but merely as the circle where $2\sigma_\theta - \sigma_r = 0$. For both Tresca's and von Mises' criteria $\partial\sigma_r/\partial r$ is zero ($\partial\sigma_\theta/\partial r$ being finite), and $\sigma_\theta - \sigma_r$ is equal to k , just within the inner boundary of the rigid annulus. It follows from the equation of equilibrium that $\partial h/\partial r$ is equal to $-hp/2c$ just within the radius c/ρ ; thus, the slope of the surface changes discontinuously. It may be shown from the Lévy-Mises equation that $\partial\sigma_\theta/\partial r$ has the value $-3kp/2c$ just within the radius c/ρ , and from the incompressibility equation that $\partial v/\partial r$ has the value $-1/2c(\partial h/\partial c = -1/\rho \partial h/\partial r$ at this point).

Since it is clearly immaterial whether the pressure at any radius is applied by an external agency or through the displacement of an inner annulus of the plate, the stress and velocity in an element depend only on what happens beyond this radius. Expressed mathematically, the equations are hyperbolic and the solution is determined by the boundary conditions on the line $r=c$ in the (r, c) plane. Moreover, since the plate is infinite the stress and velocity are functions only of the relative distance from the plastic boundary; that is, they are functions of the single parameter r/c . In particular, the stress distribution around a hole expanded from some finite radius is identical with that in the corresponding annulus in a plate in which a hole has been enlarged from zero radius and in which, therefore, the distribution of stress remains geometrically

similar*. It is the latter problem that has been investigated by Taylor, using Tresca's criterion. Let $\theta=r/c$, $\sigma=-\sigma_\theta/2k$ ($0 \leq \sigma \leq \frac{1}{2}$), and put $\sigma_r=-2k$ in (17). Then

$$\frac{\partial}{\partial r} = \frac{1}{c} \frac{d}{d\theta}, \quad \frac{\partial}{\partial c} = -\frac{\theta}{c} \frac{d}{d\theta};$$

$$\frac{\theta h'}{h} = \sigma - 1, \quad (v - \theta) \frac{h'}{h} + v' + \frac{v}{\theta} = 0, \quad v' = \left(\frac{2 - \sigma}{2\sigma - 1} \right) \frac{v}{\theta},$$

. . . (18)

where a dash denotes differentiation with respect to θ .

Eliminating h'/h :

$$v' = \frac{(2 - \sigma)(\sigma - 1)}{2(\sigma^2 - \sigma + 1)}, \quad \frac{v}{\theta} = \frac{(2\sigma - 1)(\sigma - 1)}{2(\sigma^2 - \sigma + 1)}. \quad . . . (19)$$

Eliminating v :

$$\theta \frac{d}{d\theta} \left[\frac{(2\sigma - 1)(\sigma - 1)}{(\sigma^2 - \sigma + 1)} \right] = -\frac{3(\sigma - 1)^2}{(\sigma^2 - \sigma + 1)}.$$

Integrating, and using the boundary condition $\sigma = \frac{1}{2}$ when $\theta = 1/\rho$:

$$\ln \left(\frac{1}{\rho\theta} \right) = -\frac{1}{3} \left(\frac{1 - 2\sigma}{1 - \sigma} \right) + \frac{1}{2} \ln \left\{ \frac{3(1 - \sigma)^2}{(\sigma^2 - \sigma + 1)} \right\} + \frac{1}{\sqrt{3}} \tan^{-1} \left(\frac{1 - 2\sigma}{\sqrt{3}} \right).$$

. . . (20)

Now the yield criterion $\sigma_r = -2k$ is valid only when σ_θ is negative. As θ decreases, σ decreases steadily from $\frac{1}{2}$ and becomes zero when $\theta = 1/\rho'$, where

$$\ln \left(\frac{\rho'}{\rho} \right) = -\frac{1}{3} + \frac{1}{2} \ln 3 + \frac{\pi}{6\sqrt{3}} = 0.5183; \quad \rho' = 2.769.$$

. . . (21)

The value for ρ' obtained by Taylor, by approximate numerical integration of equations equivalent to (18), agrees very closely with (21). It is not difficult to show that

$$\frac{h}{h_0} = \left[\frac{1}{2(1 - \sigma)} \right]^{\frac{1}{3}} \exp \left[\frac{2}{\sqrt{3}} \tan^{-1} \left(\frac{1 - 2\sigma}{\sqrt{3}} \right) \right], \quad . . . (22)$$

where h_0 is the initial thickness of the plate. The thickness at the radius where σ_θ is zero is thus

$$h = h_0 \left(\frac{1}{2} \right)^{\frac{1}{3}} \exp \left(\frac{\pi}{3\sqrt{3}} \right) = 1.453 h_0. \quad . . . (23)$$

Equations (19), (20) and (22) give v and h/h_0 as functions of $\theta = r/c$, parametrically in terms of σ . The displacement u is most conveniently

* If the plate is *finite*, the stress distribution is the same in the plastically *deforming* material provided the entire plate is not plastic; the breadth of the rigid plastic annulus is, of course, different.

calculated from the equation of incompressibility in the form

$$h_0 s \, ds = h r \, dr \quad (c \text{ constant}),$$

where $s=r-u$ is the initial radius to an element.

Integrating:

$$\frac{c^2}{\rho^2} - s^2 = \frac{c^2}{\rho^2} - (r-u)^2 = 2c^2 \int_{\theta}^{1/c} \frac{h\theta}{h_0} d\theta. \quad . \quad . \quad . \quad (24)$$

Alternatively, we may note that the displacement must be of the form

$$u(r, c) = cF(\theta).$$

Hence

$$\frac{\partial u}{\partial r} = F', \quad \frac{\partial u}{\partial c} = F - \theta F',$$

and

$$v = \frac{Du}{Dc} = \left(\frac{\partial}{\partial c} + v \frac{\partial}{\partial r} \right) u = (v - \theta)F' + F; \quad F = 0 \text{ when } \theta = 1/\rho.$$

$v(\theta)$ having been determined, this equation serves to calculate $F(\theta)$.

When θ is less than $1/\rho'$, that is, when r is less than $0.361c$, σ_θ becomes positive again and the yield criterion reverts to $\sigma_\theta - \sigma_r = 2k$. The corresponding modification of (18) is such that an analytic solution appears to be impossible and the equations must be integrated numerically.

TABLE I.

$\theta=r/c$	$\sigma_r/2k$	$\sigma_\theta/2k$	v	h/h_0	u/c
0.361	-1.000	0.0	0.180	1.453	0.053
0.350	-0.997	0.003	0.191	1.503	0.061
0.340	-0.988	0.012	0.202	1.561	0.070
0.330	-0.973	0.027	0.214	1.636	0.081
0.320	-0.948	0.052	0.227	1.734	0.094
0.310	-0.910	0.090	0.240	1.868	0.111
0.300	-0.855	0.145	0.254	2.065	0.133
0.295	-0.819	0.181	0.261	2.201	—
0.290	-0.772	0.228	0.268	2.388	0.169
0.285	-0.705	0.295	0.275	2.684	—
0.2805	-0.500	0.500	0.2805	3.835	0.2805

It is here that Taylor's results are seriously in error, due to an insufficiently accurate method of integration. The method adopted here was to replace each first derivative by a simple finite difference across a small interval in θ and to take mean values of other quantities. The resulting equations were solved by a trial-and-error process for each interval, until consistent values were obtained for the unknowns. The results are given in Table I.

These values are thought to be accurate to 1 per cent or better. Since the derivatives σ'_r and h' become infinite at the edge of a hole expanded

from zero radius it was necessary to use an analytic series solution near the edge. It may be shown that

$$\frac{v-\alpha}{\theta-\alpha} = -1 + A(\theta-\alpha)^{\frac{1}{2}} - \frac{34A^2}{15}(\theta-\alpha)^{\frac{3}{2}} + \dots,$$
$$\frac{\sigma_r}{2k} = -\frac{1}{2} + A(\theta-\alpha)^{\frac{1}{2}} - \frac{13A^2}{6}(\theta-\alpha)^{\frac{3}{2}} + \dots,$$

Fig. 3.

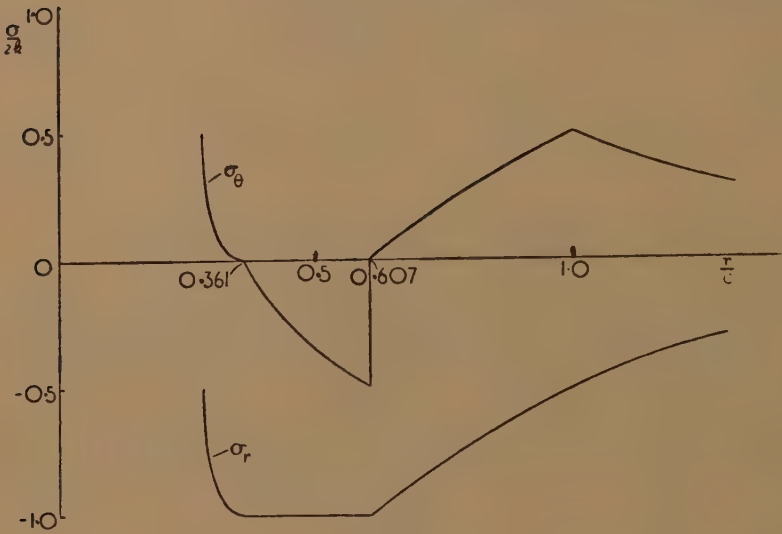
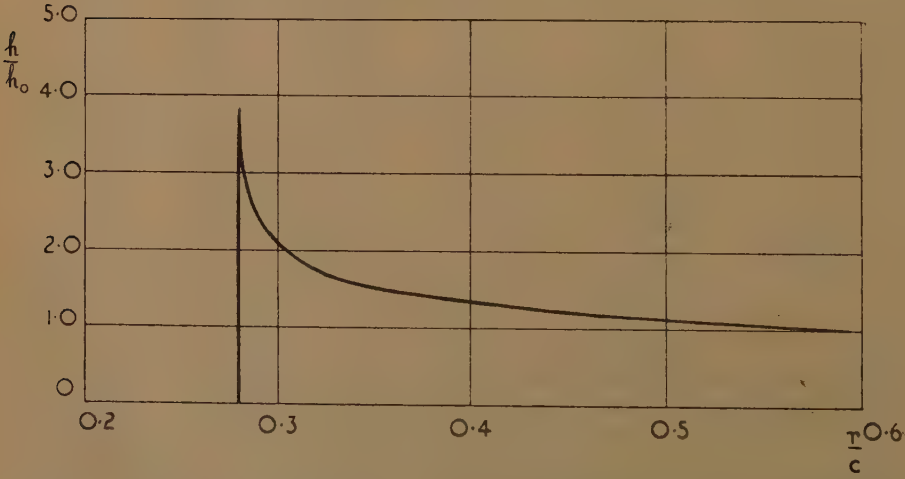


Fig. 4.



where α is the least value of θ . These equations may be solved for A and α , using the values of v and σ_r obtained at the point $\theta=0.285$ by numerical integration. It is found that $A=-0.929$, $\alpha=0.2805$. The

series expansions can be applied as far as $\theta=0.295$ with negligible error in v and about 2 per cent error in σ_r .

The distribution of stress is shown in fig. 3, and the contour of the deformed surface in fig. 4. At the edge of the hole, h/h_0 has the value 3.835, as against Taylor's value of 2.66. The almost exact agreement of the present value with that found experimentally by Taylor (see the photograph in his paper) is certainly fortuitous, since the theory can only be regarded as approximate near the edge of the hole where the coronet is knife-edged. Also, the experimental material was lead, the plastic behaviour of which departs considerably from that contemplated in the analysis. It should be mentioned that the experiments indicated that the symmetrical mode of deformation may be unstable; once the hole reached a critical size in relation to the plate thickness, the configuration became unsymmetrical and the plate bent out of its plane.

ACKNOWLEDGMENT.

I am indebted to the Director of the British Iron and Steel Research Association for permission to publish this paper.

REFERENCES.

- MANDEL, J., 1942, *Equilibres par tranches planes des solides* (Paris : Louis-Jean).
SOKOLOVSKY, W. W., 1945, *Prikladnaia Matematika i Mekhanika*, **9**, 111.
TAYLOR, G. I., 1948, *Quart. J. Mech. and App. Maths.*, **1**, 103.
SOKOLOVSKY, W. W., 1946, *Theory of Plasticity* (Moscow), 222.
SOKOLOVSKY, W. W., *ibid.*, p. 229.
HILL, R., 1949, *Quart. J. Mech. and App. Maths.*, **2**, 40.
NADAI, A., 1931, *Plasticity* (New York and London : McGraw-Hill Book Co.), p. 191.

XCII. *On the Theory of the Magnetic Double Refraction in Non-polar Liquids.*

By OLLE SNELLMAN, Fil.dr.,

Institute of Physical Chemistry, Upsala, Sweden *.

[Received July 15, 1949.]

SUMMARY.

A simplified theory of the double refraction in non-polar liquids is given. It is proposed that Langevin's theory for gases must be corrected by two terms, one for the asymmetry of the polarization cavity, and another for the potential barrier round the molecule. Putting in these terms gives results that agree rather well with data for benzene. Data for more substances are needed.

* Communicated by the Author.

For a long time there has been a lively discussion on the polarization phenomena in dielectrics. To a large extent this discussion has dealt with polar liquids, while the relations in non-polar liquids have been considered to be fairly well clarified.

For this, one has virtually merely discussed the Clausius-Mosotti polarization formula $(\epsilon - 1)/(\epsilon + 2) = P/v$. This equation, is, however, quite insensitive for small changes in the structure of the liquids owing to the relatively simple formation of averages by which it is derived. Thus, a transition from the gas to the liquid phase gives an alteration of the molecular polarization of benzene of only 4 per cent, whilst the Cotton-Mouton constant changes by a factor of about 2. This sensitivity to the structural state is further shown by nitrobenzene, which, when newly distilled, has a considerably higher Cotton-Mouton Constant than that for nitrobenzene which has been allowed to age for a few hours and has assumed the final structure of the liquid (Snellman 1944).

However, as yet, there is not enough experimental material for a discussion; the values for the magnetic double refraction in gases especially are lacking. The author will give here only a qualitative theoretical treatment of the problem for non-polar liquids with reference to the surface of an anisotropic interaction cavity and a potential barrier. A more detailed analysis of the problem cannot be made before the question of the effect of the density on the molecular parameters has been solved. This density effect does not yet seem to have been subjected to closer treatment. That such an effect exists is shown by the experiments of Salant and others (Breit and Salant, 1930).

THE LOCAL FIELD.

For gases the local field which acts on a molecule can be written, according to Lorentz (1909) as

$$F = E_0 + (4/3)\pi\kappa E_0,$$

where E_0 is the external field and κ is the electrical susceptibility of the medium.

The phenomenological description which is the basis of the Lorentz derivation of this formula uses the conception of a spherical cavity with the considered molecule as centre. The effect of the surface charges of the cavity, arising from the polarization of the medium, on the field at the centre of the cavity is then calculated with the help of the theorem of Gauss. This imaginary cavity must be large enough to contain so many molecules that they can be considered as a continuum but in the optical field, regarded quasi static, it must not be more than a fraction of the wavelength of the light. The effect of the molecules inside the sphere is neglected. This latter can only be done if the molecules are symmetrical or arbitrarily distributed.

Many objections have been raised against this conception. As early as 1920 Lundblad pointed out that the influence of nearby oscillators must be considered. One must assume that the influence of these is

small and the artificial character of the method then becomes clear. Lundblad suggested an interaction cavity that is not a sphere, or, on the other hand, an asymmetric charge distribution on the sphere so that the influence of nearby oscillators is compensated. This basis of explanation alone, as Lundblad assumed, ought not, however, to be correct. We shall return to this later.

Kirkwood (1939-40) and van Vleck (1937) have also discussed the great artificialities in the use of the Lorentz field when it is applied to dipole molecules.

However, with non-polar molecules, the Lorentz field holds surprisingly well for the calculation of polarization. Van Vleck (1937) has also shown that it holds statistically when the polarized molecules can be considered as a collection of quantized harmonic oscillators in an array of cubic symmetry. Why it should be so good when applied to polarization formula is difficult to say, because, as van Vleck remarks, such a consideration can scarcely correspond to reality. Kirkwood (1936), by a statistical method, has also found that the Lorentz field is the right one for non-polar molecules, although certain corrections are needed.

Kirkwood, in his treatment, corrects the Lorentz field, partly with respect to what he calls rotational fluctuations and partly with respect to translational fluctuations. The latter fluctuations originate from term fluctuations so that the average value of the sixth power of the inverse distance between two molecules is not the same as the square of the average cube of the inverse distance. This correction which takes account of centrally symmetrical forces in a gas has also been used by Kirkwood for liquids. Whether this extension is allowable without further considerations cannot, however, be said. We prefer to include this effect more generally in a potential barrier.

Kirkwood's rotational function corresponds to the aspherical interaction volume in our method of treatment.

The following derivation of the double refraction phenomena is based on the Langevin orientation theory. Debye (1934) has made a complete derivation of this theory with consideration to the striction problem, for the case of the Lorentz field for a spherical cavity. The striction terms disappear, however, when the double refractions (and not the absolute refractive indices) are calculated, and hence we shall not include this term in our calculations. Moreover it can be said that the assumptions of Debye only hold for ideal gases. It must be considered a mistake that on the whole he discusses his formula for liquids.

THE PROBLEM OF ORIENTATION.

The anisotropic interaction volume alone cannot give a description of double refraction phenomena in liquids, as one soon finds. The discrepancy is particularly striking if one compares the temperature dependence for the different double refractions. One must allow for the effect of the intermolecular forces which cause a certain correlation to exist between the orientation of the neighbouring molecules. These

intermolecular forces which are probably of many different kinds can be introduced, at least for non-polar molecules, as a potential barrier (W), Debye (1935) and Fowler (1935). In the general case this is a very complicated function of the molecular orientation. A more formal treatment of the problem can be made, however, if W is assumed to depend on the angle between the axes of a pair of molecules. Assuming the potential to be a symmetric function we can develop W (according to Mueller (1936)) as a function of the angle between a considered molecule and the direction of the intermolecular forces. In the following we shall also consider the molecules as having rotational symmetry.

For further calculations we must establish certain coordinate systems. We first imagine a coordinate system x, y, z , fixed in space with the z -axis in the direction of the external field and the x and y axes perpendicular to it and to each other. Further, a coordinate system 1, 2, 3, is fixed to the molecule, with the 1-axis in the direction of the major axis of the considered molecule. A transformation from the one coordinate system to the other can be expressed for the simplest case, the rotationally symmetrical, with the angles θ and ψ (fig. 1).

Fig. 1.

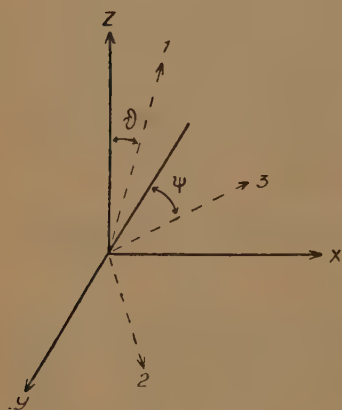
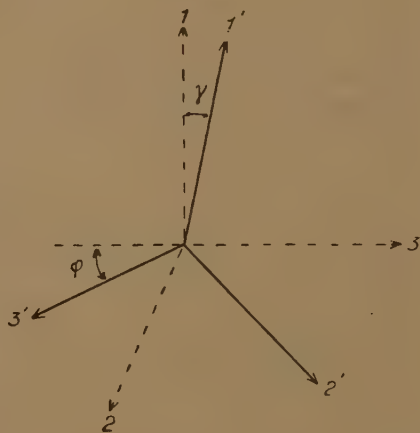


Fig. 2.



	x	y	z
1	$\sin \theta \cos \psi$	$\sin \theta \sin \psi$	$\cos \theta$
2	$-\cos \theta \cos \psi$	$-\cos \theta \sin \psi$	$\sin \theta$
3	$\sin \psi$	$-\cos \psi$	0

Similarly, we can give the potential barrier a coordinate system $1', 2', 3'$, where $1'$ lies parallel to the direction of the axis of the intermolecular forces. The angles between this system and the coordinate system 1, 2, 3, are γ and ϕ respectively. In this case one obtains (fig. 2)

	1	2	3
$1'$	$\cos \gamma$	$\sin \gamma \cos \phi$	$\sin \gamma \sin \phi$
$2'$	$\sin \gamma$	$-\cos \gamma \cos \phi$	$-\cos \gamma \sin \phi$
$3'$	0	$\sin \phi$	$-\cos \phi$

We can expand the potential barrier W in the series

$$-W = W_0 + W_1 \cos^2 \gamma, \quad \dots \quad (1.1)$$

where W_0 is a normalizing factor and W_1 is the energy required to turn the molecule against the intermolecular forces.

This potential barrier affects the distribution of the molecules. The effect on the distribution can be written as

$$dN_{(\gamma, \phi)} = A \exp(W_1 \cos^2 \gamma / kT) \sin \gamma d\gamma d\phi. \quad \dots \quad (1.2)$$

Normalized,

$$dN_{(\gamma, \phi)} = \exp\left(\frac{W_1 \cos^2 \gamma}{kT}\right) \sin \gamma d\gamma d\phi / \left\{ 2\pi \int_0^{\frac{1}{2}\pi} \exp\left(\frac{W_1 \cos^2 \gamma}{kT}\right) \sin \gamma d\gamma \right\}. \quad \dots \quad (1.3)$$

If an external magnetic field is applied the change in the potential energy for a molecule can be written as

$$u = -\frac{1}{2} \{ \kappa_1 H_1^2 + \kappa_2 (H_2^2 + H_3^2) \}, \quad \dots \quad (1.4)$$

where κ_i is the magnetic susceptibility of the molecule in the direction i . The field H_j acts along the axes $1'$, $2'$, $3'$. Since the changes due to the magnetic polarization can be completely neglected, we can write:

$$\left. \begin{aligned} H_1 &= H(\cos \theta \cos \gamma + \sin \theta \sin \gamma \cos \phi), \\ H_2 &= H(\cos \theta \sin \gamma - \sin \theta \cos \gamma \cos \phi), \\ H_3 &= H \sin \theta \sin \phi. \end{aligned} \right\} \quad \dots \quad (1.5)$$

Inserting (1.5) in (1.4) gives

$$u = -\frac{1}{2} H^2 \{ \kappa_1 + (\kappa_1 - \kappa_2)(\cos^2 \theta \cos^2 \gamma + \sin^2 \theta \sin^2 \gamma \cos^2 \phi + 2 \cos \theta \sin \theta \cos \gamma \sin \gamma \cos \phi) \}. \quad \dots \quad (1.6)$$

For the distribution when the field is applied one gets

$$dN = \frac{\exp(W_1 \cos^2 \gamma / kT) \sin \gamma d\gamma d\phi}{2\pi \int_0^{\frac{1}{2}\pi} \exp\left(\frac{W_1 \cos^2 \gamma}{kT}\right) \sin \gamma d\gamma} \cdot A' \exp(-u/kT) \sin \theta d\theta d\phi. \quad \dots \quad (1.7)$$

u/kT is a small quantity and a series expansion is therefore permissible

$$dN = \frac{\exp(W_1 \cos^2 \gamma / kT) \sin \gamma d\gamma d\phi}{2\pi \int_0^{\frac{1}{2}\pi} \exp\left(\frac{W_1 \cos^2 \gamma}{kT}\right) \sin \gamma d\gamma} \cdot A' \left\{ 1 + \frac{H^2(\kappa_1 - \kappa_2)}{2kT} \cdot (\cos^2 \theta \cos^2 \gamma + \sin^2 \theta \sin^2 \gamma \cos^2 \phi + 2 \cos \theta \sin \theta \cos \gamma \sin \gamma \cos \phi) \right\} \sin \theta d\theta d\phi, \quad (1.8)$$

or normalized

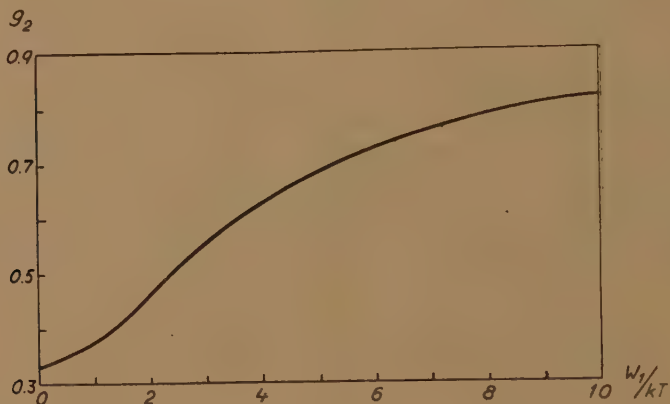
$$dN = \frac{\exp(W_1 \cos^2 \gamma / kT) \sin \gamma d\gamma d\phi}{4\pi^2 \int_0^{\frac{1}{2}\pi} \exp\left(\frac{W_1 \cos^2 \gamma}{kT}\right) \sin \gamma d\gamma} \cdot \left\{ 1 + \frac{H^2(\kappa_1 - \kappa_2)}{2kT} \cdot [\cos^2 \theta (\cos^2 \gamma - g_2) + \sin^2 \theta (\sin^2 \gamma \cos^2 \phi - \frac{1}{2} \{1 - g_2\}) + 2 \cos \theta \sin \theta \cos \gamma \sin \gamma \cos \phi] \right\} \times \sin \theta d\theta d\phi. \quad \dots \quad (1.9)$$

According to Mueller (1936) we have introduced the following function into (1.9):

$$g_2 = \int_0^{\frac{1}{2}\pi} \exp\left(\frac{W_1 \cos^2 \gamma}{kT}\right) \cos^2 \gamma \sin \gamma d\gamma / \int_0^{\frac{1}{2}\pi} \exp\left(\frac{W_1 \cos^2 \gamma}{kT}\right) \sin \gamma d\gamma. \quad (1.10)$$

Fig. 3 shows the variation of g_2 with W_1/kT according to Mueller. If $W_1=0$ we have free rotation and $g_2=\frac{1}{3}$, if W_1 is very great we have $g_2=1$. g_2 can consequently vary between the limits $\frac{1}{3}$ and 1.

Fig. 3.



THE OPTICAL THEORY.

We will at first regard the case of an optical aspheric interaction cavity in a homogeneous liquid. We can then introduce a constant ν in the expression for the local field instead of the value $4\pi/3$

$$\mathbf{F} = (1 + \nu\kappa)\mathbf{E}_0. \quad (2.1)$$

The factor ν depends on the shape of the interaction cavity. If the interaction cavity has rotational symmetry we have for the two different principal directions of the cavity

$$\left. \begin{aligned} \nu_1 &= \frac{4}{3}\pi - 2 \cdot \frac{4}{3}\pi s_2, \\ \nu_2 &= \frac{4}{3}\pi + \frac{4}{3}\pi s_2, \end{aligned} \right\} \quad (2.2)$$

and

$$s_1 + 2s_2 = 1,$$

where s_2 depends only on the eccentricity of the interaction cavity according to Maxwell (1881).

Thus the inner field becomes a function of the direction. For the moments in the different directions we have

$$\left. \begin{aligned} m_1 &= b_1(1 + \frac{4}{3}\pi\kappa)(1 - 2 \cdot \frac{4}{3}\pi s_2 \{ \kappa / (1 + \frac{4}{3}\pi\kappa) \}) \mathbf{E}_0, \\ m_2 &= b_2(1 + \frac{4}{3}\pi\kappa)(1 + \frac{4}{3}\pi s_2 \{ \kappa / (1 + \frac{4}{3}\pi\kappa) \}) \mathbf{E}_0. \end{aligned} \right\} \quad (2.3)$$

Since we disregard the field of the molecules surrounding the considered molecule, we can lay our cavity so that it only encloses the free space for the regarded molecule, and s_2 is then determined by this space. If we regard the inner field as isotropic, we can correct the axis, and then we get for the main polarizabilities of this molecule

$$\left. \begin{aligned} \bar{b}_1 &= b_1(1 - 2 \cdot \frac{4}{3} \pi s_2 \{ \kappa / (1 + \frac{4}{3} \pi \kappa) \}), \\ \bar{b}_2 &= b_2(1 + \frac{4}{3} \pi s_2 \{ \kappa / (1 + \frac{4}{3} \pi \kappa) \}). \end{aligned} \right\} \dots \dots \dots (2.4)$$

We will now treat our case with a potential barrier round this molecule. It is possible to lay an ellipsoidal cavity round the molecule which has the property that the atoms within give no contributions to the inner field. Let this cavity have the form factor L_2 . In the same manner as before the inner field will be anisotropic and then we have

$$\left. \begin{aligned} F_1 &= (1 + \frac{4}{3} \pi \kappa)(1 - 2 \cdot \frac{4}{3} \pi L_2 \{ \kappa / (1 + \frac{4}{3} \pi \kappa) \}) E_0 = h_1 E_0, \\ F_2 &= (1 + \frac{4}{3} \pi \kappa)(1 + \frac{4}{3} \pi L_2 \{ \kappa / (1 + \frac{4}{3} \pi \kappa) \}) E_0 = h_2 E_0, \end{aligned} \right\} \dots \dots (2.5)$$

h_1 and h_2 are introduced here as proportionality factors between the internal and external fields and, as before,

$$L_1 + 2L_2 = 1. \dots \dots \dots (2.6)$$

If we then allow a beam of light to be incident on a doubly refracting medium along the y -axis and if the electric vector has the magnitude E_0 , the field will act along the locally preferred directions and the moments in these directions become with the help of the coordinate system given above

$$\left. \begin{aligned} P_{1'} &= \bar{b}_1 F_{1'}, \\ P_{2'} &= \bar{b}_2 F_{2'}, \\ P_{3'} &= \bar{b}_2 F_{3'}, \end{aligned} \right\} \dots \dots \dots (2.7)$$

where $F_{i'}$ is the magnitude of the internal field in the direction i' .

If now the incident field is assumed to be parallel to the z -axis, its projections on the axes 1, 2, 3, are

$$\left. \begin{aligned} F_1 &= (E_0)_z h_1 \cos \theta, \\ F_2 &= (E_0)_z h_2 \sin \theta, \\ F_3 &= 0, \end{aligned} \right\} \dots \dots \dots (2.8)$$

since we have the ellipsoidal interaction cavity (*cf.* equation (2.5)).

Furthermore, the projections of this field on the drawn directions are

$$\left. \begin{aligned} F_{1'} &= F_1 \cos \gamma - F_2 \sin \gamma \cos \phi, \\ F_{2'} &= F_1 \sin \gamma - F_2 \cos \gamma \cos \phi, \\ F_{3'} &= F_2 \sin \phi. \end{aligned} \right\} \dots \dots \dots (2.9)$$

The relative moment along the z -axis then becomes

$$\begin{aligned} p_z &= p_{1'}(\cos \theta \cos \gamma + \sin \theta \sin \gamma \cos \phi) \\ &\quad + p_{2'}(\cos \theta \sin \gamma - \sin \theta \cos \gamma \cos \phi) + p_{3'} \sin \theta \sin \phi. \end{aligned} \quad (2.10)$$

Combining (2.10) with (2.7), (2.8), (2.9) gives

$$p_z = (\bar{b}_1 - \bar{b}_2)(E_0)_z \{ h_1 \cos^2 \theta \cos^2 \gamma + h_2 \sin^2 \theta \sin^2 \gamma \cos^2 \phi \\ + (h_1 + h_2)(\cos \theta \sin \theta \cos \gamma \sin \gamma \cos \phi) \} + \bar{b}_2(E_0)_z (h_1 \cos^2 \theta + h_2 \sin^2 \theta). \quad (2.11)$$

For the case where the electric vector is parallel to the x -axis we have the equations

$$\left. \begin{aligned} p_{1'} &= \bar{b}_1 F_{1'} \\ p_{2'} &= \bar{b}_2 F_{2'} \\ p_{3'} &= \bar{b}_2 F_{3'} \end{aligned} \right\} \left. \begin{aligned} F_1 &= (E_0)_x h_1 \sin \theta \cos \psi \\ F_2 &= -(E_0)_x h_2 \cos \theta \cos \psi \\ F_3 &= (E_0)_x h_2 \sin \psi \end{aligned} \right\} \left. \begin{aligned} F_{1'} &= F_1 \cos \gamma + F_2 \sin \gamma \cos \phi \\ &\quad + F_3 \sin \gamma \sin \phi \\ F_{2'} &= F_1 \sin \gamma - F_2 \cos \gamma \cos \phi \\ &\quad - F_3 \cos \gamma \sin \phi \\ F_{3'} &= F_2 \sin \phi - F_3 \cos \phi \end{aligned} \right\}$$

and for the moment along the x -axis

$$p_x = p_{1'}(\sin \theta \cos \gamma \cos \psi - \cos \theta \sin \gamma \cos \psi \cos \phi + \sin \gamma \sin \psi \sin \phi) \\ + p_{2'}(\sin \theta \sin \gamma \cos \psi + \cos \theta \cos \gamma \cos \psi \cos \phi - \cos \gamma \sin \psi \sin \phi) \\ + p_{3'}(-\cos \theta \cos \psi \sin \phi - \sin \psi \cos \phi) \quad (2.12)$$

or

$$p_x = (E_0)_x (\bar{b}_1 - \bar{b}_2) \{ -h_1 \cos^2 \theta \cos^2 \gamma \cos^2 \psi - h_2 \sin^2 \theta \cos^2 \psi \cos^2 \phi \\ - (h_1 + h_2) \cos \theta \sin \theta \cos \gamma \sin \gamma \cos^2 \psi \cos^2 \phi \\ + (h_1 + h_2) \sin \theta \cos \gamma \sin \gamma \cos \psi \sin \psi \sin \phi \\ + h_1 \cos^2 \gamma \cos^2 \psi + h_2 \sin^2 \gamma (\cos^2 \psi \cos^2 \phi + \sin^2 \psi \sin^2 \phi) \\ - 2h_2 \cos \theta \sin^2 \gamma \cos \psi \sin \psi \cos \phi \sin \phi \} \\ + (E_0)_x \bar{b}_2 (h_1 \sin^2 \theta \cos^2 \psi + h_2 \cos^2 \theta \cos^2 \psi + h_2 \sin^2 \psi). \quad (2.13)$$

We can now form the mean values for the moments in the different directions

$$\left. \begin{aligned} p_z &= \int p_z dN, \\ p_x &= \int p_x dN. \end{aligned} \right\} \quad (2.14)$$

We can further define the mean polarizability $\bar{\alpha}$ as $P = \bar{\alpha} E_0$ and thus have for the difference

$$P_z / (E_0)_z - P_x / (E_0)_x = (\bar{\alpha}_z - \bar{\alpha}_x). \quad (2.15)$$

Substituting the values of the different quantities and carrying out the integration we have

$$\bar{\alpha}_z - \bar{\alpha}_x = \frac{1}{30} \frac{H^2}{2kT} (\kappa_1 - \kappa_2) (\bar{b}_1 - \bar{b}_2) \{ 2h_1(g_2 - g_2^2) + h_2(1 - g_2^2) \}. \quad (2.16)$$

If we regard $\bar{\alpha}$ as the apparent polarization in the Lorentz-Lorentz dispersion formula

$$(n^2-1)/(n^2+2) = \frac{4}{3}\pi N' \bar{\alpha},$$

we can use this formula directly. Thus we have for the difference in the directions z and x

$$6n(n_z - n_x)/(n^2+2)^2 = \frac{4}{3}\pi N'(\bar{\alpha}_z - \bar{\alpha}_x), \quad \dots \quad (2.17)$$

n is the refractive index and N' the number of molecules per unit volume.

From equations (2.16) and (2.17) we get

$$n_z - n_x = \{(n^2-1)(n^2+2)/4n\} \{H^2(\kappa_1 - \kappa_2)/kT\} \{6(\bar{b}_1 - \bar{b}_2)/180\} \\ \times \{1 + 2g_2 - 3g_2^2 + L_2(1 - 4g_2 + 3g_2^2)(n^2-1)/(n^2+2)\} \dots \quad (2.18)$$

For magnetic double refraction the following relation has been experimentally confirmed

$$n_z - n_x = \lambda C H^2, \quad \dots \quad (2.19)$$

λ is the wave-length of the light, C is a constant for the substance in question (Cotton-Mouton Constant).

For a substance obeying the ideal gas laws the Cotton-Mouton Constant can be calculated, from the assumption of rotational symmetry of the molecule, to be

$$C = \{(n^2-1)(n^2+2)/4n\lambda\} \{(\kappa_1 - \kappa_2)2(b_1 - b_2)/45\alpha kT\}. \quad \dots \quad (2.20)$$

In order to compare the different constants more easily we want to use the Molecular Cotton-Mouton Constant. This is defined by Briegleb as

$$C_M = C^* \{6n^2/(n^2+2)^2\} (M/\rho), \quad \dots \quad (2.21)$$

C^* is a constant defined by $C^* = C\lambda/n$. Thus we have for the case of a gas

$$C_M^V = 4N(\kappa_1 - \kappa_2)(b_1 - b_2)/45kT, \quad \dots \quad (2.22)$$

where N is Avogadro's number.

According to equation (2.6) we can write

$$\bar{b}_1 - \bar{b}_2 = (b_1 - b_2) \{1 - [L_2(2b_1 + b_2)/(b_1 - b_2)][(n^2-1)/(n^2+2)]\}. \quad (2.23)$$

Introducing this relation in equation (2.18) gives the Cotton-Mouton Constant as

$$C = \{(n^2-1)(n^2+2)/4n\lambda\} \{(\kappa_1 - \kappa_2)2(b_1 - b_2)/45\alpha kT\} \\ \times [1 - s_2(2b_1 + b_2)(n^2-1)/(b_1 - b_2)(n^2+2)] \\ \times \frac{3}{4}[1 + 2g_2 - 3g_2^2 + L_2(1 - 4g_2 + 3g_2^2)(n^2-1)/(n^2+2)]. \quad \dots \quad (2.24)$$

If (2.24) is compared with (2.20)

$$C_M^L/C_M^V = \frac{3}{4}[1 - s_2(2b_1 + b_2)(n^2-1)/(b_1 - b_2)(n^2+2)] \\ \times [1 + 2g_2 - 3g_2^2 + L_2(1 - 4g_2 + 3g_2^2)(n^2-1)/(n^2+2)]. \quad \dots \quad (2.25)$$

We will now, as a first approximation, regard the two form factors as similar, $s_2=L_2$. It is seen that the contribution of the environment of the molecule depends on the shape of the molecule. The form factors s_2 and L_2 have nothing to do with the size of the ellipsoid, only with the shape. We thus obtain as a first approximation

$$\begin{aligned} C_M^L/C_M^V = & \frac{3}{4}[1+2g_2-3g_2^2+L_2\{1-4g_2+3g_2^2-(1+2g_2-3g_2^2) \\ & \times (2b_1+b_2)/(b_1-b_2)\}(n^2-1)/(n^2+2)]. \quad . \quad . \quad . \quad (2.26) \end{aligned}$$

This is the term by which the gas value of C_M must be corrected in order to obtain the liquid value. The temperature function is essentially due to g_2 .

In the case where the interaction volume is spherical we have ($L_2=0$)

$$C_M^L/C_M^V = \frac{3}{4}(1+2g_2-3g_2^2). \quad . \quad . \quad . \quad (2.27)$$

If we have no potential barrier $g_2=\frac{1}{3}$ ($L_2 \neq 0$)

$$C_M^L/C_M^V = 1 - L_2(2b_1+b_2)(n^2-1)/(b_1-b_2)(n^2+2). \quad . \quad . \quad (2.28)$$

For completely hindered rotation ($g_2=1$) we obtain, in contrast to Mueller, the ratio $C_M^L/C_M^V=0$, i. e. no influence on applying the field. This is to be expected, and not a finite value. Mueller has not considered the distribution to change when one introduces the potential barrier.

From magnetic double refraction alone one cannot calculate g_2 and L_2 . For the Kerr Effect one can easily calculate a similar expression. We then get for the Kerr Constant (if $\epsilon=n^2$)

$$\begin{aligned} K_M^L/K_M^V = & \frac{3}{4}[1+2g_2-3g_2^2+L_2\{3(1-4g_2+3g_2^2) \\ & -2(1+2g_2-3g_2^2)(2b_1+b_2)/(b_1-b_2)\}(n^2-1)/(n^2+2)]. \quad . \quad (2.29) \end{aligned}$$

If we had $L_2=0$ we should have $C_M^L/C_M^V=K_M^L/K_M^V$. Such is not the case for the substance (benzene) where all the quantities are measured. Furthermore, one would have, if $W_1=0$, that $g_2=\frac{1}{3}$, and

$$(K_M^L/K_M^V-1)/(C_M^L/C_M^V-1)=2,$$

which is not the case. Obviously one needs to consider both factors.

By graphical interpolation it is possible to determine g_2 and L_2 from (2.26) and (2.29).

One can verify the magnitudes of the values obtained in the polarization formula. From the theory given one finds, with

$$R=\{(n^2-1)/(n^2+2)\}(M/\rho),$$

$$R^L/R^V=1-(1+3g_2)L_2(b_1-b_2)(n^2-1)/(b_1+2b_2)(n^2+2). \quad . \quad (2.30)$$

The difference between the liquid and gas values is, however, so small that only qualitative verification can be done in this way.

Sufficient data to calculate g_2 and L_2 and hence verify the theory have been obtained only for benzene.

For benzene one finds from the literature the following data at 20° C.

$$C_M^L/C_M^V=0.490, \quad K_M^L/K_M^V=0.309, \quad R^L/R^V=0.963, \quad (13, 14, 15, 16)$$

$$\text{and} \quad b_1=63.5 \times 10^{-25}, \quad b_2=123.1 \times 10^{-25}.$$

By means of graphical interpolation the following values at 20° C. are obtained from C_M^L/C_M^V and K_M^L/K_M^V ,

$$g_2=0.705, \quad L_2=(n^2-1)/(n^2+1)=-0.750.$$

From these values one gets $W_1=5.5kT$, and $L_2=-0.26$.

If the value of g_2 is introduced into the expression for R^L/R^V , one gets $L_2=-0.22$.

The value of W_1 seems to be of a reasonable order of magnitude. The values of L_2 from the two different expressions agree satisfactorily within the limits of experimental errors. If further corrections for the influence of the density on the polarizability constant are necessary, one has to expect several terms in the factor by which $L_2(n^2-1)/(n^2+2)$ must be multiplied. In order to investigate how far this simplified theory is valid, however, data for substances other than benzene are needed.

The author wishes to express his gratitude to Mr. E. J. Burge for helping him check the equations.

REFERENCES.

- BREIT, G., and SALANT, E. O., 1930, *Phys. Rev.*, **36**, 871.
 DEBYE, P., 1935, *Physik. Zeit.*, **36**, 100, 193.
 DEBYE, P., and SACK, H., 1934, *Handb. der Radiologie*, Bd. 6 : 2. (Leipzig.)
 FOWLER, R. H., 1935, *Proc. Roy. Soc. A*, **149**, 1.
 KIRKWOOD, J., 1936, *J. Chem. Phys.*, **4**, 592 ; 1939, *Ibid.* **7**, 911 ; 1940, *Ann. of N.Y. Acad. of Sciences*, **40**, 315.
 KÖNIG, H., 1938, *Ann. Phys.*, **31**, 289.
 LORENTZ, H. A., 1909, *The Theory of Electrons*. (Leipzig.)
 LUNDBLAD, R., 1920, *Diss., Upsala*.
 MAXWELL, J. C., 1881, *A Treatise on Electricity and Magnetism*. (Oxford.)
 MUELLER, M., 1936, *Phys. Rev.*, **50**, 547.
 MATULL, E., 1934, *Ann. Phys.*, **21**, 345.
 SNELLMAN, O., 1944, *Diss., Upsala*.
 STUART, H. A., and VOLKMANN, H., 1933, *Zs. f. Phys.*, **83**, 444.
 VAN VLECK, J. H., 1937, *J. Chem. Phys.*, **5**, 556.
 WASASTJERNA, J., 1924, *Soc. Sc. Fenn. Comm. Phys. Math.*, **2**, 9.

XIII. *The Magnetic Double Refraction of Liquid Mixtures—*
 I. *Benzene in Various Solvents.*

By EDWARD J. BURGE and OLLE SNELLMAN, Fil.dr.,
 The Institute of Physical Chemistry, University of Upsala, Sweden*.

[Received July 15, 1949]

ABSTRACT.

The theories for non-polar substances, and the previous experimental studies of the magnetic double refraction of liquid mixtures are reviewed. Measurements are given for benzene in the following solvents: carbon tetrachloride, cyclohexane, pentane, hexane, heptane, and dodecane. The value of the Cotton-Mouton constant of dodecane is given for the first time, $C_m = -2.81 \times 10^{-14}$, and compared with the values for other paraffins. The results are discussed in relation to the theories of Snellman and Ramanadham. Complete agreement is found with Ramanadham's theory for the simple case of benzene in carbon tetrachloride. The values also support Snellman's theory, but more data are required for a closer check, especially values for vapours.

INTRODUCTION.

COTTON and Mouton (1905) discovered that when nitrobenzene is placed in a magnetic field it exhibits a difference of refractive index between the directions parallel and perpendicular to the field. Methods of measurement depend on the analysis of elliptically polarized light obtained by passing a beam of light, plane polarized at 45° to the azimuth, through the liquid in a direction perpendicular to the applied magnetic field.

Most of the studies of magnetic double refraction (henceforth abbreviated to MDR; also known as Cotton-Mouton effect, and magnetic birefringence) have been made with pure liquids in order to investigate the relation between MDR and chemical structure. The studies have continued although a complete theory is available only for the state of an ideal gas. Much remains to be done before a rigorous theory can be developed. The purpose of the present investigation is to study the MDR of mixtures of benzene and organic solvents, with special reference to the molecular constant at infinite dilution.

* Communicated by the Authors.

THEORIES FOR NON-POLAR SUBSTANCES.

The Langevin-Born theory considers the molecules to have electric and magnetic anisotropies represented by ellipsoids of revolution, and also allows for permanent electric and magnetic moments. It uses a Boltzmann distribution of the molecular orientation and is only applicable to the state of an ideal gas, since the Lorentz field is used without corrections. (See *e.g.* Marx, *Handbuch der Radiologie*, 6, 745 (1925)).

Some theoretical investigators (*e.g.* Fowler and Guggenheim, *Statistical Thermodynamics* (Cambridge 1939)) consider that the Lorentz field can be applied almost directly to non-polar fluids. Such a view will give fairly good results for dielectric measurements where second-order effects in the interaction between the molecules do not appear. However, all the experimental investigations show that the application of the L.-L. dispersion formulæ and the L.-B. theory to the MDR of liquid mixtures does not give correct results.

Ramanadham (1936) has given a theory for the MDR of liquid mixtures. He applies a correction to the L.-B. theory which allows for the variation of the anisotropy of the polarization field when the percentage of the solute is varied. He takes the observed values of the refractive indices of the two mixtures and calculates variable parameters of the polarization field according to the theory of Raman and Krishnan (1927 a, b). Ramanadham states that the values obtained are in general agreement with the experimental data of Chinchalkar (1933) (excluding the data for nitrobenzene), and claims that a quantitative description of the MDR in liquid mixtures is given for the first time. It is therefore of interest to make a critical check of the experimental values of MDR as given by Chinchalkar, and the values of the refractive indices given and used by Ramanadham. This has been done and will be reported in a later paper. It appears that the claim of good agreement is not well founded.

From the theoretical point of view also, Ramanadham's theory does not seem to be correct, for it assumes the problem to be too simple. He considers that only the optical polarization field is altered when two liquids are mixed, thus providing only one variable to be adjusted. This does not seem to be the general case with MDR. One of us, in another paper (Snellman 1949), has been able to show that for pure benzene, which is a non-polar substance, agreement can only be obtained in the results from the comparison of measurements of MDR, Electrical Double Refraction (EDR), and Molar Refractivity (R), if at least two independent variables are used.

This new theory can also, of course, be applied to the case of liquid mixtures. Details are to be found in the paper mentioned above. Here it is sufficient to state that the ratio of the molecular constant of a substance at infinite dilution to the molecular constant for the gas (or vapour) can be expressed in terms of two variables; the anisotropy of the inner field, L, and a potential barrier contained in the factor g_2 .

This potential barrier allows for the rotational and translational movements of the molecules in the cavity considered.

We obtain

$$\{C_M^S\}_0/C_M^V = 3/4 \left[(1 + 2g_2 + 3g_2^2) + L_2 \frac{n^2 - 1}{n^2 + 2} \left\{ 1 - 4g_2 + 3g_2^2 - (1 + 2g_2 - 3g_2^2) \frac{2b_1 + b_2}{b_1 - b_2} \right\} \right] \quad (1)$$

For the case of a liquid at infinite dilution, n will be the refractive index of the solvent. For the solute molecule (at a given concentration in the solvent considered) b_1 and b_2 will be the optical polarization constants. The terms g_2 and L_2 will vary with the solvent. If the theory is correct, we can obtain information about the structure of the solvent from the magnitude of the factor g_2 . In order to test the theory and to obtain values for closer study, it is necessary to obtain data for the EDR (Kerr effect), and for Molar Refractivity. The number of data available for discussion along the lines mentioned above is very small, and the data are often inaccurate. Different kinds of measurement must be made for more rigorous checking of present theories and data, in order, eventually, to form a more complete picture of the fine structure of liquids and the phenomena of polarization. As a contribution to the need for more data, we have begun to study the MDR of non-polar substances in non-polar liquids. Furthermore, efforts are being made to measure the MDR of certain organic vapours and gases. These measurements are urgently needed and still only available for benzene and, to a lower degree of accuracy, for nitro-benzene, König (1938).

EXPERIMENTAL STUDIES OF OTHER INVESTIGATORS.

The most comprehensive work so far has been carried out by Chinchalkar (1933). He showed that the MDR is rarely directly proportional to the concentration and can be greater or less than that expected from a simple linear relation or from the Lorentz equation for the refractive index of mixtures, in combination with Langevin's theory of MDR. The number of observations for each mixture is small, and the accuracy is low because of the compensator used. Chinchalkar (1935) also studied solutions of certain aromatic hydrocarbons, containing two or more benzene rings, in carbon tetrachloride and other solvents. Measurements at only one concentration for each mixture have been made by Mahajan (1936) for some aromatic hydrocarbons in organic solvents, using the same experimental arrangement as Chinchalkar.

Solutions of nitrobenzene in carbon tetrachloride were first studied by Cotton and Mouton (1913), later by Chinchalkar (1933), and most completely by Piekara (1934) who also studied the variation with temperature from 10° C. to 40° C. Piekara and Goldet (1934) treated the special case of nitrobenzene in hexane and the variation of its MDR with temperature. The critical concentration and temperature, previously found in measurements of dielectric polarization, were clearly demonstrated.

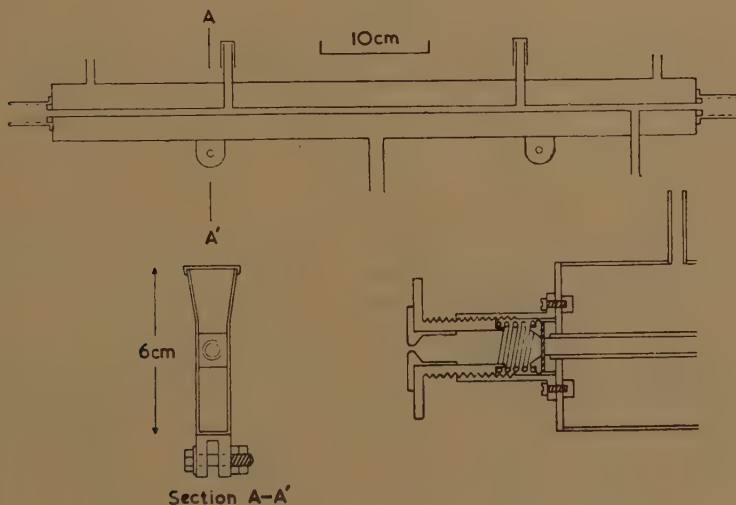
The measurements of Szivessy and Richartz (1928) are difficult to interpret because they are mostly made with polar substances in polar solvents. The measured points are few, and in some cases extrapolation to infinite dilution is very uncertain.

APPARATUS.

For the measurement of MDR this consists of the electromagnet, the tube containing the liquid to be investigated, and the optical arrangements.

(i) *The electromagnet.* A description of the electromagnet is given by Björnsthål (1934), and Dreyfus (1934). Snellman (1944) has made some measurements of the fields for polepieces of various shapes, different pole-gaps, and currents up to 1700 A.* The polepieces used here are wedge-shaped with surfaces 58×1 cm.² the pole-gap being 14 mm. The current was kept constant at 1500 A., giving a field of about 40,000 Oersteds. The magnet was used with the axis of the coils in a horizontal position.

Fig. 1.



Tube for the liquid, and detail of window attachment.

(ii) *The tube for the liquid.* This is simply a brass tube, gilt on the inside, 60 cm. long and 6 mm. internal diameter, with a surrounding water jacket (fig. 1). The windows are 18 mm. diameter microscope slides, selected to have minimal double refraction, and held in position by the arrangement shown in the figure. A thin film of a solution of sucrose in glycerine is placed between the window and the flange of the tube. The slightest convections inside the tube deflect the beam of light (c. 1.5×1.5 mm.²). In practice the use of a thermostat in conjunction with the water jacket was found to be a disadvantage, probably because

* As these papers are not readily available, a general description is to be published in the *Journal of Scientific Instruments*.

the system of entrance and exits was too simple. With water in the water-jacket, but no circulation, equilibrium at room temperature is obtained after about an hour—depending on the nature of the liquid in the tube. The room temperature was observed to be fairly constant at 18.5°C . ($\pm 0.5^{\circ}\text{C}$).

(iii) *Optical arrangements.* These are identical with those of Snellman (1944) except for the addition of a circular diaphragm (D_2) at each end of the tube (fig. 1). The light path used is that designed by Björnståhl (1939) and is in four groups as follows:—(a) mercury arc, (b) condensing lens, slit of prismatic monochromator set for 5461 Å , exit slit of monochromator (S), adjustable square diaphragm (D_1), and lens (L), (c) polarizing nicol prism, 2 mm. diam. circular diaphragm (D_2) at the entrance to the tube containing the liquid, and circular diaphragm at the exit—2 or 3 mm., and (d) half-shade plate (H), compensator plate, analysing nicol prism, and telescope. The images of S and D_1 , formed by the lens L, are made to fall at D_2 and H, respectively.

ADJUSTMENTS.

Each of the groups of components (a), (b) and (d), is clamped to an optical bench supported by tables that are adjustable in a vertical direction and mounted on wheels which can be jacked up off the floor. In the case of group (b), however, the condensing lens is clamped to one optical bench and D_1 and L are on another, since the entrance and exit beams of the monochromator are not in the same direction. Group (c) is mounted on the optical bench attached to the magnet. With the exception of the monochromator, all components can be raised and lowered, rotated about a vertical axis, and translated horizontally along two directions mutually at right angles, of which one is the direction of the light beam.

In the absence of the tube, the position of the light beam is made as central as possible using a slip of brass, 14 mm. wide, drilled perpendicular to its face with three holes of 1, 2 and 3 mm. diameter respectively, on the centre line of its long direction. This slip slides between the pole surfaces, and the adjustment is effected by means of D_1 and L. The tube is then filled with carbon tetrachloride and placed in position with the polarizer at approximately 45° to the azimuth. The field is applied and the magnet rotated on its turntable until no Faraday effect is detectable, *i. e.* the light beam is perpendicular to the field direction. The tube is next filled with benzene, which has a large MDR, and the nicol prisms adjusted so that no phase difference is observable when the field is switched on. This indicates that the plane of polarization is parallel or perpendicular to the field direction. The two nicols are then rotated through 45° in the same direction so that they remain crossed.

DETERMINATION OF REFRACTIVE INDEX.

At a later stage of the investigation it was found to be of interest to measure the refractive index of each solution studied. This was done with an Abbé refractometer from Zeiss (Jena), calibrated for Na light.

using methylene iodide as a check on the calibration. Unless otherwise stated, the temperature for the determination of refractive indices was 18.6°C ., held constant to $\pm 0.01^{\circ}\text{C}$. by a thermostat. Only the values for the solute and solvents could be given for mixtures reported here owing to insufficient quantities of the particular samples being available for determinations at the various concentrations.

DEFINITIONS AND SYMBOLS USED.

C_m is the Cotton-Mouton Constant of a pure substance, as defined by

$$\Delta = C_m l H^2 = \delta / \lambda = 1(n_p - n_s) / \lambda, \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Δ =phase difference, l =optical path length, H =applied magnetic field, δ =difference of optical path length, λ =wavelength of light used, n =refractive index, p and s refer to directions parallel and perpendicular to the azimuth respectively.

C^* is conveniently defined by

$$C^* = C_m \lambda / n = \frac{n_p - n_s}{n} / H^2, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

n =refractive index of solution before the field is applied.

C_M is the Molecular Cotton-Mouton Constant as defined by König (1938), and is analogous to that for the Kerr constant given by Briegleb (1931, 1932) and Otterbein (1934)

$$C_M = C^* \frac{6n^2}{(n^2 + 2)^2} \cdot \frac{M}{\rho}, \quad . \quad . \quad . \quad . \quad . \quad (4)$$

M is the molecular weight, ρ is the density.

This is for the special case of a diamagnetic substance with small magnetic susceptibility. C_M is so defined that it ought to be independent of the state of aggregation, density, and refractive index. C_M^L , C_M^L , C_M^V , C_M^V , C_M^S , C_M^S , refer to the liquid and vapour states and solutions. The value of C_M^S depends on concentration, and for a comparison of theories it is convenient to determine the value obtained by extrapolation to infinite dilution, $(C_M^S)_0$. This is obtained by the extrapolation of the plot of concentration against

$$C_M^L \cdot \Delta_c / \Delta_{LT},$$

where Δ_c =observed value of Δ (corrected for the MDR of the solvent), and Δ_{LT} =value of Δ given by a linear theory of the dependence of MDR on concentration, *i. e.*

$$(C_M^S)_x = (C_M^L) \left/ \frac{(100-x)}{100} \right. + (C_M^L)_2 \frac{x}{100}, \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where x is the percentage of solute, and the suffices 1 and 2 refer to the solute and solvent respectively. The basis of this method of plotting and extrapolation is the Lorentz-Lorentz law of the additivity of refractivities,

$$R_{\text{mixture}} = \sum (R_i x_i / 100), \quad . \quad . \quad . \quad . \quad . \quad (6)$$

where i refers to the different components.

For one component we have

$$R_p - R_s = C_M^L \cdot H^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where p and s have the same meaning as before. This equation is the definition of C_M . From this it follows that

$$C_M^S = \Sigma(C_M^L)_i x_i / 100. \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

For the simple case of a binary mixture, it is clear that the method of plotting given above will give the true value of $(C_M^S)_0$, but that the shape of the curve from 0 to 100 per cent is not a representation of the variation of C_M^S with concentration.

MEASUREMENTS.

The apparatus was calibrated by means of benzene (Schering-Kahlbaum pro analysi, $n=1.5030$) using a Soleil compensator with a half-shade plate. This calibration was checked several times during the course of the experiments reported here. The value $C_m = 7.12 \times 10^{-13}$ (Snellman 1944) was considered to be the most accurate for the calculations. Each measurement was repeated three times or more during a period of several hours, sometimes overnight.

(i) Benzene in Carbon Tetrachloride.

The sample of carbon tetrachloride used (Albright and Wilson, sample 1, $n=1.4620$) showed a small negative MDR, ($C_m = -3.8 \times 10^{-15}$), considered to be unimportant for the present studies. (Allowance was made for the value of the MDR of the solvent when plotting the results). Other samples also gave small but definite values, possibly due to traces (up to 0.7 per cent) of carbon disulphide (a common impurity, with the largest negative value of MDR, $C_m = -6 \times 10^{-13}$). The values, with the angles of rotation of the Brace compensator plate, are given below. The accepted value of n at 18°C . is 1.46455.

Albright and Wilson, sample 1, $C_m = -3.8 \times 10^{-15}$, ($\theta = -22'$), $n=1.4620$.

Albright and Wilson, sample 2, $C_m = -4.0 \times 10^{-15}$, ($\theta = -23'$), $n=1.4620$.

Schering, $C_m = -4.2 \times 10^{-15}$, ($\theta = -24'$), $n=1.4620$.

"Chemically pure" (Kebo or Grave?),

$C_m = -4.2 \times 10^{-15}$, ($\theta = -24'$), $n=1.4619$.

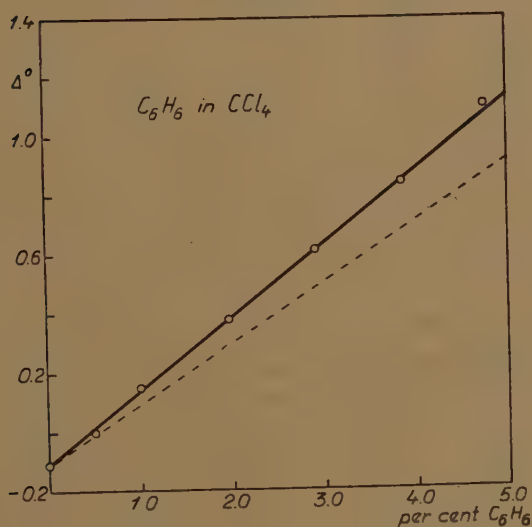
In this connection it is interesting to notice that, since the average of at least five settings of the Brace compensator, with the value of Δ_0 (phase shift of the compensator plate) $= 8^\circ 38'$, is repeatable to $1'$, the sensitivity of the present apparatus for small values of θ can be placed at $\pm 1'$. This is equivalent to a value of C_m of $\pm 2 \times 10^{-16}$, which is to be compared with the statement of Matull (1934) that C_m for carbon tetrachloride is less than $\pm 3 \times 10^{-15}$, presumably representing the sensitivity of his apparatus. This value of Matull is seen to be a little less than the values

found for the C_m of the carbon tetrachloride used here. The sensitivity of the apparatus is a maximum for small values of MDR (or, better, θ) and decreases for the larger values because the whole field of the half-shade becomes lighter, making accurate settings impossible.

In order to allow for the value of C_m for the solvent when calculating the ratio of " $\Delta_{\text{corrected}}$ " to " Δ_{LT} ", the MDR contributed by the solvent is considered to be proportional to concentration. Thus, if x is the percentage of benzene, Δ_0 the measured MDR of the solvent expressed in degrees, and Δ_{obs} the observed MDR of the x per cent mixture in degrees,

$$\Delta_c = \Delta_{\text{obs}} - \frac{(100-x)}{100} \Delta_0. \quad (9)$$

Fig. 2.

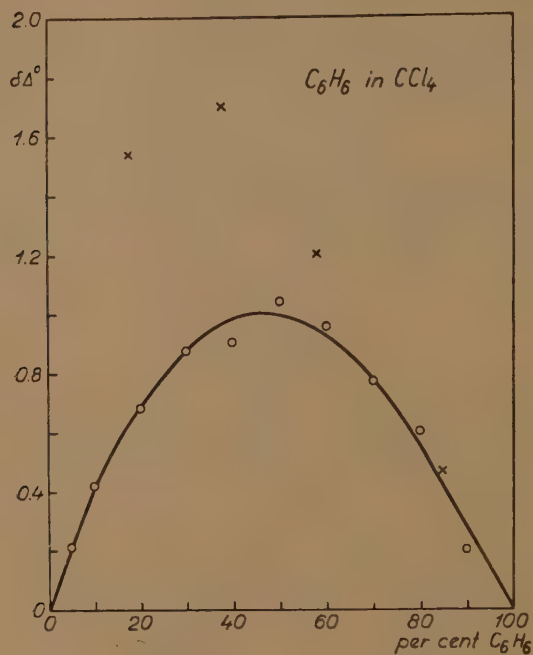


This correction is usually important for $x < 15$, but is small for higher concentrations.

The results of Chinchalkar (1933), and König (1938), suggested that measurements of low concentrations (0–10 per cent) would be most interesting.

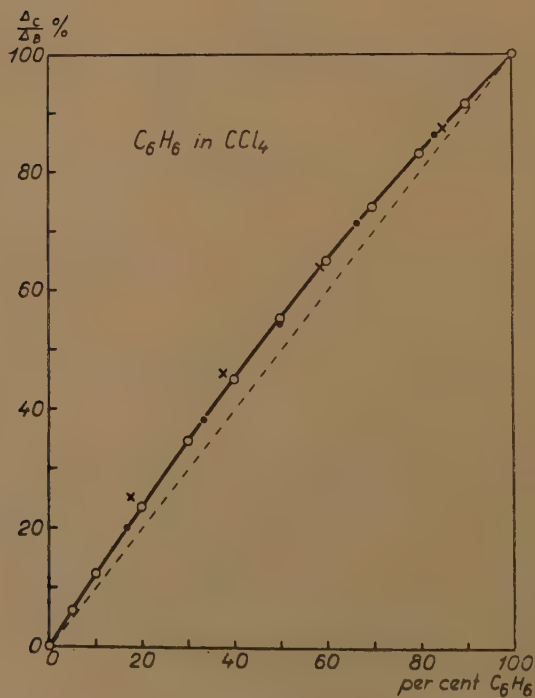
The values of Δ° (the measured MDR for the given conditions of tube length, field strength, etc., expressed in degrees) for 0–5 per cent benzene lie on a straight line (fig. 2), indicating a constant value of $C_M^S = 207 \times 10^{-17}$. The dotted line is that of the simple linear theory. Fig. 3 gives a useful method for observing discrepancies, being the variation of $\delta\Delta^\circ (= \Delta_c^\circ - \Delta_{\text{LT}}^\circ)$ with concentration. The conventional method of plotting is shown in fig. 4. The values of C_M^S for 0–100 per cent benzene are given in Table I. and fig. 5.

Fig. 3.



○ Authors. × Chinehalkar.

Fig. 4.



○ Authors. × Chinehalkar. ● Ramanadham's.

TABLE I.
Benzene in Carbon Tetrachloride.

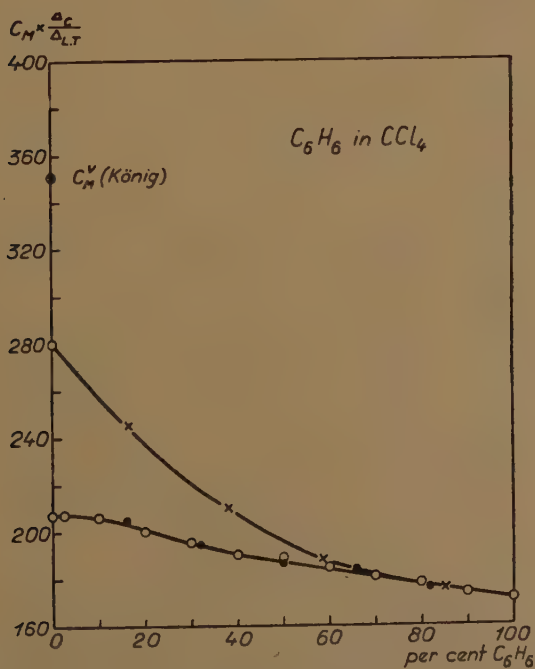
Compensator	C ₆ H ₆ %	θ	Δ°	Δ_{LT}°	Δ_c°	$C_M^L \cdot \Delta_c / \Delta_{LT}$
B ₁	0.0	— 22'	—0.11	0.0	0.0	207.2
B ₁	0.498	— 2'	—0.01	0.101	0.12	—
B ₁	0.99	+ 29'	+0.145	0.202	0.255	—
B ₁	1.96	1° 16'	0.38	0.400	0.49	—
B ₁	2.91	2° 2'	0.61	0.594	0.72	—
B ₁	3.85	2° 48'	0.84	0.786	0.95	—
B ₁	4.76	3° 40'	1.10	0.972	1.20	—
B ₂	10.0	3° 15'	2.37	2.04	2.46	206.3
B ₂	20.0	6° 30'	4.68	4.08	4.76	199.8
B ₂	30.0	9° 43'	6.92	6.12	6.99	195.2
B ₂	40.0	12° 49'	9.0	8.16	9.06	190.0
S	50.0	120	11.19	10.20	11.24	188.7
S	60.0	141	13.15	12.24	13.19	184.4
S	70.0	161	15.02	14.28	15.05	180.4
S	80.0	181	16.90	16.32	16.92	177.5
S	90.0	199	18.55	18.36	18.56	173.3
S	100.0	219	20.40	20.40	20.40	171.2

B₁=Brace, $\Delta_0=8^\circ 38'$.

B₂=Brace, $\Delta_0=21^\circ 18'$.

S=Soleil with half-shade.

Fig. 5.



○ Authors.

× Chinchalkar.

● Ramanadham's theory.

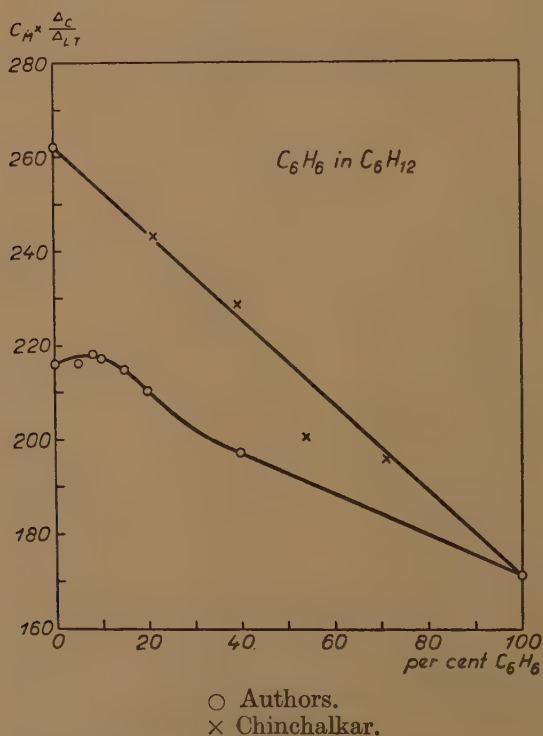
(ii) *Benzene in Cyclohexane.*

The MDR of the sample of cyclohexane used (Kebo) was $C_m = -1.47 \times 10^{-14}$ which is in good agreement with that measured by Schérer ($C_m = 1.48 \times 10^{-14}$) in terms of benzene, if Snellman's value for pure nitrobenzene is used. The results are given in Table II., and fig. 6.

TABLE II.
Benzene in Cyclohexane.

Compensator	$C_6H_6\%$	θ	Δ°	Δ_{LT}°	Δ_c°	$C_M^L \cdot \Delta_c / \Delta_{LT}$
B ₁	0.0	$-1^\circ 24'$	-0.420	0.0	0.0	216.0
B ₁	5.0	$+2^\circ 58'$	$+0.89$	1.02	1.29	216.2
B ₂	8.1	$2^\circ 22'$	1.713	1.653	2.104	218.0
B ₂	10.0	$3^\circ 3'$	2.21	2.04	2.59	217.4
B ₂	15.0	$4^\circ 48'$	3.483	3.06	3.84	214.9
B ₂	20.0	$6^\circ 29'$	4.67	4.08	5.01	210.2
B ₂	40.0	$13^\circ 5'$	9.17	8.16	9.42	197.2

Fig. 6.

(iii) *Benzene in Hydrocarbons of the Paraffin Series.*

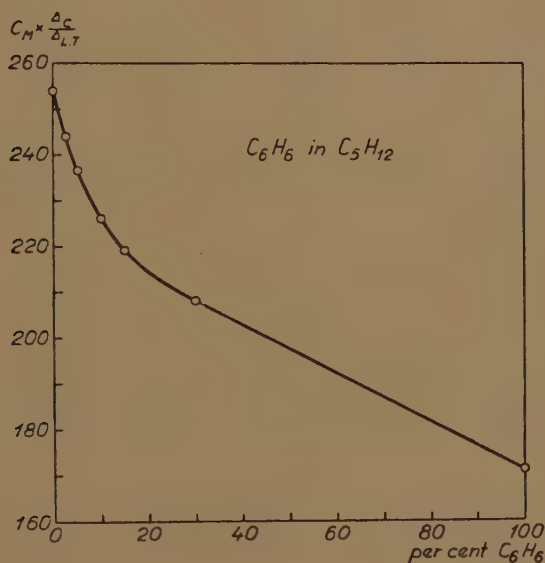
(a) Pentane. From petroleum, Grave. $n=1.3570$, accepted value $n=1.3570$. $C_m = -0.63 \times 10^{-14}$. Results in Table III., and fig. 7. The formation of bubbles at the ends of the tube made measurements difficult at small concentrations of benzene.

TABLE III.

Benzene in Pentane.

Compensator	C ₆ H ₆ %	θ	Δ°	Δ_{LT}	Δ_c°	$C_M^L \cdot \Delta_c / \Delta_{LT}$
B ₁	0.0	— 36'	—0.18	0.0	0.0	254.0
B ₁	2.5	+1° 51'	+0.552	0.51	0.727	244.0
B ₂	5.0	1° 42'	1.24	1.02	1.41	236.6
B ₂	10.0	3° 29'	2.53	2.04	2.69	226.0
B ₂	15.0	5° 13'	3.77	3.06	3.92	219.1
B ₂	30.0	10° 18'	7.33	6.12	7.45	208.0

Fig. 7.



(b) Hexane. Kebo. $C_m = -0.76 \times 10^{-14}$. Results in Table IV., and fig. 8. Still trouble with bubbles, but only slight.

TABLE IV.

Benzene in Hexane.

Compensator	C ₆ H ₆ %	θ	Δ°	Δ_{LT}	Δ_c°	$C_M^L \cdot \Delta_c / \Delta_{LT}$
B ₁	0.0	— 44'	—0.22	0.0	0.0	239.0
B ₁	5.0	+3° 54'	+1.17	1.02	1.379	231.3
B ₂	10.0	3° 25'	2.48	2.04	2.68	224.8
B ₂	30.0	10° 18'	7.33	6.12	7.48	209.4

Fig. 8.

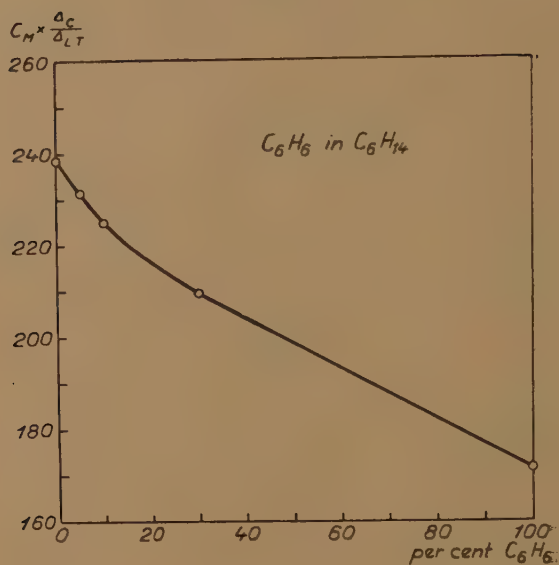
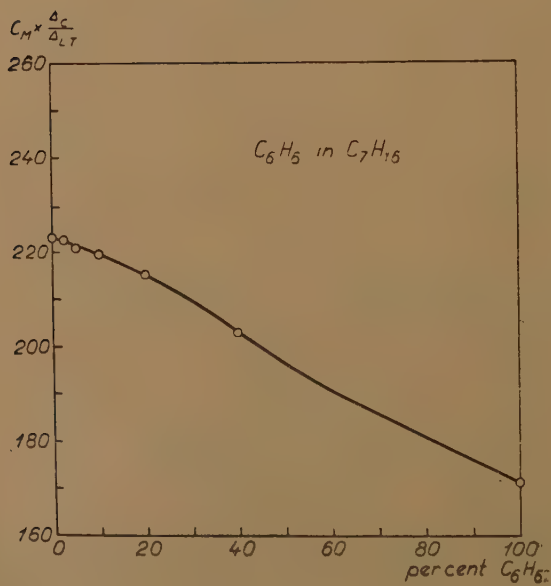


Fig. 9.



(c) Heptane. A pure sample was not available and various values were obtained for the samples tested:—

Eastman Kodak, sample 1,	$C_m = +2.10 \times 10^{-14}$,
Eastman Kodak, sample 2,	$C_m = +2.27 \times 10^{-14}$, $n = 1.4048$.
Kebo,	$C_m = -0.56 \times 10^{-14}$,
May and Baker,	$C_m = -1.01 \times 10^{-14}$, $n = 1.4015$.
Schering,	$C_m = -1.05 \times 10^{-14}$, $n = 1.4087$.

TABLE V.

Benzene in Heptane.

Compensator	$C_6H_6\%$	θ	Δ°	Δ_{LT}°	Δ_c°	$C_M^L \cdot \Delta_c / \Delta_{LT}$
B ₂	0.0	— 24'	—0.30	0.0	0.0	223.0
B ₂	2.5	+ 30'	+0.375	0.51	0.365	222.6
B ₂	5.0	1° 25'	1.03	1.02	1.315	220.8
B ₂	10.0	3° 14'	2.35	2.04	2.62	219.5
B ₂	20.0	6° 48'	4.90	4.08	5.14	215.2
B ₂	40.0	13° 35'	9.50	8.16	9.68	202.9

The accepted value for the refractive index is 1.3878 at 20° C., for $\rho = 0.6842$. The sample from Schering was used for the measurements with benzene, the results being shown in Table V., and fig. 9.

(d) Dodecane. The sample used was from Johnson, Sweden, and claimed to contain 95 per cent normal isomer. Its refractive index was 1.4232 at 18.6° C., which is to be compared with the value 1.4218 at 20° C. ($\rho = 0.7495$), given in Landolt-Börnstein, III., 1689. No previous measurement of the MDR of dodecane is to be found in the literature. The value

TABLE VI.

Benzene in Dodecane.

Compensator	$C_6H_6\%$	θ	Δ°	Δ_{LT}°	Δ_c°	$C_M^L \cdot \Delta_c / \Delta_{LT}$
B ₁	0.0	—2° 40'	—0.805	0.0	0.0	247.0
B ₁	5.0	+2° 21'	+0.705	1.02	1.465	245.7
B ₂	10.0	3° 2'	2.20	2.04	2.92	244.8
B ₂	15.0	5° 0'	3.62	3.06	4.30	240.5
B ₂	20.0	6° 50'	4.92	4.08	5.56	232.8
B ₂	40.0	13° 57'	9.62	8.16	10.15	212.6

found for this sample was $C_m = -2.81 \times 10^{-14}$. Mixtures with benzene, especially the low percentages, required rather a long time to become homogeneous owing to the syrupy nature of the dodecane. The results are given in Table VI., and fig. 10.

Fig. 10.

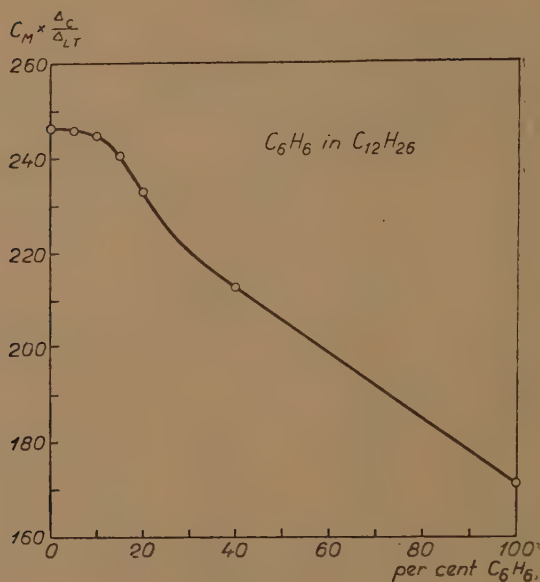
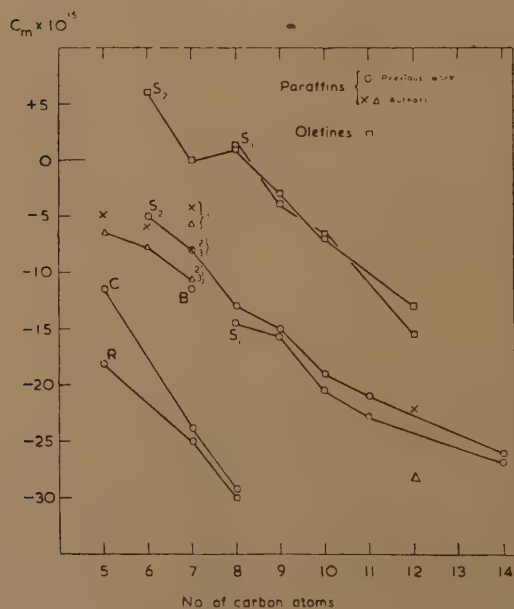


Fig. 11.



Cotton-Mouton Constant of the Paraffins and Olefines.

S_1 , Schérer (1931); S_2 , Schérer (1934).

C, Chinchalkar (1931); R, Ramanadham (1929).

B, Boorse (1934).

\times , Calculated with $C_m = 24.12 \times 10^{-13} (C_6H_5NO_3)$.

\triangle , Calculated with $C_m = 7.12 \times 10^{-13} (C_6H_6)$.

For heptane: 1, Kebo; 2, May and Baker; 3, Schering.

MDR OF PARAFFINS AND OLEFINS.

In order that the values of the MDR of the paraffins may be compared with previous determinations, a graph of the variation of MDR with number of carbon atoms is given in fig. 11. Presumably the values (S_2) given by Schérer (1934) are the most reliable. It is seen that the new values are in close agreement with S_2 when the value for nitrobenzene is taken as $C_m = 24.12 \times 10^{-13}$, as used by Schérer (1934). Using this same value for nitrobenzene, there is not complete agreement between S_1 (Schérer, 1931) and S_2 (Schérer, 1934). The larger values of Chinchalkar (1931) and Ramanadham (1929) may be due to their using samples containing higher percentages of the normal isomer, but it is more probable that their apparatus was not sufficiently sensitive for the measurement of such small values of MDR. This point will be discussed further in connection with the values for dilute solutions of benzene given by Chinchalkar (1933).

DISCUSSION OF RESULTS.

(i) *Benzene in Carbon Tetrachloride.*

The previous measurements of Chinchalkar (1933) lie on a smooth curve (fig. 5), giving a value for $(C_M^S)_0$ of 280. Except for the higher concentrations, the present results show smaller values for C_M^S , and give $(C_M^S)_0 = 207.2$. A study of the measurements of Chinchalkar has been made by plotting them as in fig. 5. This has shown that the variation of the results is often large, and few mixtures give such a smooth curve as that for benzene in carbon tetrachloride. This can also be seen in fig. 6, for benzene in cyclohexane. A study of Chinchalkar's results for mixtures, and some measurements of pure liquids (Chinchalkar 1931, 1933) leads to the conclusion that the Rayleigh bent glass plate compensator used by him was not sufficiently accurate for small values of MDR for his magnitudes of H and l .

(ii) *Cyclohexane.*

The disagreement with Chinchalkar's measurements is again evident. It will be noticed that the lower concentrations have been measured here in order to allow extrapolation to zero concentration. Chinchalkar gives $(C_M^S)_0 = 262$. The new value is 216.

(iii) *Paraffins.*

The shapes of the curves form a sequence from an almost tangential approach at zero concentration for pentane, to a straight line for heptane, and a perpendicular approach for dodecane. The values of $(C_M^S)_0$ are: C_5H_{12} , 254; C_6H_{14} , 239; C_7H_{16} , 223; $C_{12}H_{26}$, 247; suggesting a minimum value near nonane. Unfortunately, none of the members of the series between C_7 and C_{12} was available for measurement.

The molecular constant in all cases investigated is considerably smaller than that for benzene in the vapour state, and the value for liquid benzene is not much smaller than that for benzene in mixtures. The value of $(C_M^S)_0$ varies considerably for different solvents. Thus the solvent plays an important rôle in the magnetic double refraction of even non-polar substances.

It was expected that an investigation of benzene in a homologous series would reveal regularities, but this does not seem to be the case. Even the variation of the molecular constant with dilution is not regular for the different members of the series.

A hint that the irregularities lie in the structure of the long-chain hydrocarbons is to be found in the discussion of Glasstone, Laidler and Eyring on the activation energies for viscous flow of these substances (*The Theory of Rate Processes*, p. 497: New York 1941). Departures from the theoretically predicted activation energies occur after eight carbon atoms in the chain. They consider this result to suggest that the chains no longer act as single units but in sections and that the work required to form a "hole" for the unit of flow is thus less. Such structural alterations must also be visible in the MDR. If the benzene

TABLE VII.

Molecular Constant of Benzene in Various Solvents.

Solvent	Observer	$(C_M^S)_0 \times 10^{17}$
C Cl ₄	1	280
C Cl ₄	2	207.2
C ₅ H ₁₂	2	254
C ₆ H ₁₄	2	239
C ₇ H ₆	2	223
C ₁ H ₂₆	2	247
C ₆ H ₁₂	1	262
C ₆ H ₁₂	2	216
(C H ₃) ₂ CO	1	280
C S ₂	1	192
C ₆ H ₆	Snellman	$C_M^L = 171.2$
C ₆ H ₆	König	$C_M^V = 351$

Observers; Chinchalkar—1
Authors —2

molecule can move more freely when the chains of the hydrocarbons begin to act in sections a rise of the molecular constant can be expected.

Thus there seems to be agreement between the hypothesis of higher hydrocarbons acting in sections and the fact that our values for the molecular constant pass through a minimum in these solvents at 8-9 carbon atoms.

A summary of the values of $(C_M^S)_0$ is given in Table VII.

COMPARISON OF RESULTS WITH RAMANADHAM'S THEORY.

Ramanadham (1936) gives values for MDR expected on his theory for mixtures of benzene and carbon tetrachloride, using observed values of refractive index. He also gives the figures for mixtures of toluene and carbon tetrachloride, and carbon disulphide and carbon tetrachloride.

When plotted in the form of fig. 4, excellent agreement was claimed for benzene and toluene solutions, fair agreement for carbon disulphide, and it was pointed out that no agreement was found for nitrobenzene. Studies of solutions of toluene, to be reported later, suggested that a recalculation of the values given would be of interest.

The values are given in terms of relative birefringence,

$$\frac{C_m^S}{C_m^L} = \frac{n^2 - 1}{n} \times \frac{n_1}{n_1^2 - 1} \times \frac{B_1 - B_2}{B_1^* - B_2^*} \times \frac{b_1 + 2b_2}{(b_1 + 2b_2) + 3b'\nu_2/\nu_1}, \quad (10)$$

where

n = refractive index of the mixture (observed), see *loc. cit.* p. 386.

n_1 = refractive index of the solute, given with n .

B_1, B_2 are the effective optic moments of the molecule of the birefringent component in the solution. They vary with concentration and are calculated from

$$B_1 = b_1 \left\{ 1 + \frac{p_1}{4\pi} (n_L^2 - 1) \right\}, \quad \dots \quad (11)$$

where, n_L = refractive index of the pure solute,

$\frac{p_1}{4\pi}$ = the values given on p. 392 (*loc. cit.*). The term p_1 , and for

B_2 the term p_2 , gives the anisotropy of the polarization field.

b_1, b_2 are the optic moments of the molecule in the gaseous state, assuming an axis of symmetry ($b_2 = b_3$). They are given for several liquids, including benzene and carbon disulphide but not toluene, by Ramanadham (1934).

B_1^*, B_2^* are the effective optic moments of the molecule in the pure liquid, and are given as B_1 and B_2 (sic) in the same place as b_1 and b_2 .

ν_1, ν_2 are the number of molecules per c.c. of the mixture for the respective components.

b' is the optic moment of the solvent molecule (assumed to have $b'_1 = b'_2 = b'_3$). It is calculated from the formula of Lorentz,

$$b' = \frac{n_2^2 - 1}{n_2^2 + 2} \times \frac{3}{4\pi \times \nu'}, \quad \dots \quad (12)$$

where n_2 = refractive index of the solvent.

$$\nu' = \frac{N\rho}{M} = \text{number of molecules per c.c. of solvent.}$$

The constants used and the values obtained are given in Table VIII. together with Ramanadham's values. In this case there is good agreement between the new calculations and those of Ramanadham. The expression $C_M^L(C_m^S/C_m^L)(100/x)$ is equivalent to the ordinate $(C_M^L \cdot \Delta_c/\Delta_{LT})$ of fig. 5, and the values are plotted on the same graph for comparison. It is seen that the agreement with the present observations is exact, within the limits of experimental error, whereas there is only agreement with the values of Chinchalkar for the larger percentages of benzene.

Investigations of solutions of toluene in carbon tetrachloride have shown that this agreement is not to be found in all cases. It is possible that Chinchalkar's theory holds only for simple mixtures, such as those consisting of symmetrical molecules like benzene and carbon tetrachloride.

TABLE VIII.

Ramanadham's Theory.

Benzene in Carbon Tetrachloride.

$n_{C_6H_6}=1.4960$	$n_{CCl_4}=1.4562$	$B_1^*=10.760 \times 10^{-24}$				
$\frac{n_1}{n_1-1}=1.2084$	$b'=1.040 \times 10^{-23}$	$B_2^*=16.550 \times 10^{-24}$				
$\nu_{C_6H_6}=6.778 \times 10^{21}$	$\nu_{CCl_4}=6.241 \times 10^{21}$	$b_1=6.457 \times 10^{-24}$				
$(C_M^L)_{C_6H_6}=171.2 \times 10^{-17}$		$b_2=12.790 \times 10^{-24}$				
Vol. C_6H_6 in 6 c.c.	$\frac{n^2-1}{n}$	$\frac{B_1^*-B_2^*}{B_1-B_2}$	$\frac{b+2b_2}{(b_1+2b_2)+\frac{\nu_2}{\nu_1}3b'}$	(C_m^S/C_m^L)	$C_M^L(C_m^S/C_m^L) \cdot \frac{100}{x}$	Ramanadhām.
5	0.8191	1.020	0.8480	0.8561	175.9	176.8
4	0.8107	1.060	0.6905	0.717	184.0	185.4
3	0.8007	1.065	0.5272	0.543	185.9	187.3
2	0.7910	1.105	0.3579	0.378	194.3	195.2
1	0.7808	1.161	0.1824	0.200	205.0	204.8

COMPARISON OF RESULTS WITH SNELLMAN'S THEORY.

For a discussion of Snellman's theory, values for the EDR and the polarization are needed both for the mixtures and the gaseous state.

When all these values are given, the constants can be determined from two of the equations and checked in the third one. In the formula for the polarization the term $1 = R^V/R^L$ appears. This term is small because R^V/R^L is almost unity, so that the values obtained using this equation for the evaluation of g_2 or L_2 will be unsatisfactory. A small error in R^V/R^L will cause a large error in the values for the constants obtained. It was thought that the greatest error would lie in the determination of R^V on account of the experimental difficulties encountered in measuring it. Therefore it seemed more convenient to evaluate the constants from the magnetic and electric double refractions and then to determine a value of R^V using these constants, and compare it with the experimental results.

For benzene the values given by König (gas value—g.v. 1938), Snellman (liquid value—l.v. 1944) for the MDR have been used, and for the EDR values those of Stuart (g.v. and l.v. 1934). For the main polarization axes ($b_1 = 63.5 \times 10^{-25}$; $b_2 = 123.1 \times 10^{-25}$; cf. Table VIII.) the values determined by Stuart have been used. The values of Wasastjerna (g.v. 1924) and Pesce (l.v. 1935) have been used for the optical polarization (R_n), and for the dielectric polarization (R_e) those of Höjendahl (g.v. 1928) and Niini (l.v. 1936).

For benzene in carbon tetrachloride the values of Briegleb (EDR), Ramanadham (R_n) and Niini (R_e) were used.

A determination of benzene in carbon tetrachloride at infinite dilution will give the following values : $g_1=0.570$, $L_2=-0.29$, $P_n^V=27.6$, $P_e^V=28.1$ (experimental values $P_n^V=27.2$, $P_e^V=28.0$).

In this case the agreement between the experimental values of the polarization and those obtained from the values of g and L can be considered satisfactory ; and the agreement is as good as in the case of benzene.

For the other mixtures investigated it is only possible to check the theory for hexane and heptane, because for the other solvents used values from the other measurements are still lacking. But there is another difficulty here which makes the values a little obscure ; the composition of these solvents from different chemical companies varies and thus it is difficult to compare the results with those of different investigators. Therefore the values obtained are not in good agreement.

Heptane : $g_2=0.61$, $L_2=-0.24$, $P_e^V=27.4$ (expt. $P_e^V=28.0$).

Hexane : $g_2=0.54$, $L_2=-0.27$, $P_n^V=27.0$ (expt. $P_n^V=27.2$).

There is nothing in the values which directly contradicts the theory, but on the other hand, too few values can yet be considered for a more rigorous test of the theory. It is necessary to collect much more data in these fields for the testing of theories. Preferably the data should be collected so that it is possible to investigate a few substances in several different solvents, rather than investigate a great number of substances in one or two solvents as is usually the case. It is much easier to discover the rules for the optical behaviour of the substances when the experiments are performed in the former manner rather than in the latter. Several investigators have thought it possible to obtain values for the molecular structure of the substances from data collected for different substances in the same solvents (*e. g.* Otterbein), but this is quite impossible because the structural factors are different for different substances in the same solvent.

ACKNOWLEDGMENTS.

One of us (E.J.B.) wishes to thank the Swedish Government and the Swedish Institute for financial support, which enabled him to pursue these studies in Upsala, and to Prof. The Svedberg for providing the excellent facilities of his laboratory for this work. We are indebted to Mr. Evald Hellman for help with the calculations.

REFERENCES.

- BJÖRNSTÅHL, Y., 1934, The New Laboratory of Physical Chemistry at Upsala University (Lund).; 1939, *J. Opt. Soc. America*, **29**, 201.
 BOORSE, H. A., 1934, *Phys. Rev.*, **46**, 187.
 BRIEGLEB, G., 1931, *Z. Phys. Chemie B*, **14**, 97 ; 1932, *Ibid.*, B, **16**, 249.
 CHINCHALKAR, S. W., 1931, *Ind. J. Physics*, **6**, 165 ; 1933, *Ibid.*, **7**, 491 ; 1935, *Prcc. Ind. Acad. Sci.*, **2A**, 525.
 COTTON, A., and MOUTON, H., 1905, *C.R.*, **141**, 317 ; 1913, *Ibid.*, **156**, 1456.

- DREYFUS, L., 1935, *ASEA Journal*, **12**, 8.
 FOWLER, and GUGGENHEIM, 1939, *Statistical Thermodynamics* (Cambridge).
 GLASSSTONE, LAIDLER, and EYRING, 1941, *The Theory of Rate Processes* (New York).
 GOLDET, A., 1933, *C.R.*, **197**, 1612.
 HÖJENDAHL, K., 1928, *Dissertation* (Köbenhavn).
 KÖNIG, H., 1938, *Ann. d. Physik*, **31**, 289.
 MARX, 1934, *Handbuch der Radiologie*, VI/2, p. 754 (article of Debye and Sack).
 MATULL, E., 1934, *Ann. d. Physik*, **21**, 345.
 MAHAJAN, L. D., 1936, *Phil. Mag.* (7) **22**, 717.
 NIINI, 1936, *Ann. Ac. Sc. Fenn. Ser. A*, XLVI., No. 1.
 OTTERBEIN, G., 1934, *Phys. Zeit.*, **35**, 249.
 PESCE, B., 1935, *Gazz. Chim. Ital.*, **65**, 440.
 PIEKARA, A., 1934, *J. Phys. Rad.*, **7**, 541.
 PIEKARA, A., and GOLDET, A., 1934, *C.R.*, **199**, 271.
 RAMANADHAM, M., 1929, *Ind. J. Physics*, **4**, 15; 1934, *Proc. Ind. Acad. Sci.*, **1**, 281; 1936, *Ibid.*, **3**, 384.
 RAMAN, C. V., and KRISHNAN, K. S., 1928 a, *Proc. Roy. Soc. A*, **117**, 1; 1928 b, *Ibid.* **A**, **117**, 589.
 SCHÉREER, M., 1931, *C.R.*, **192**, 1223; 1934, *Dissertation* (Paris).
 SNELLMAN, O., 1944, *Dissertation* (Upsala); 1949, *Phil. Mag.*
 SNELLMAN, O., and BURGE, E. J., 1949, *J. Scient. Instr.*
 STUART, H. A., and VOLKMANN, H., 1933, *Z. f. Physik*, **83**, 444.
 SZIVESSY, G., and RICHARTZ, M., 1928, *Ann. d. Physik*, **86**, 393.
 WASASTJERNA, J. A., 1924, *Soc. Sci. Fenn. Comm.*, **2**, No. 13.

XCIV. Investigation of Soft Radiations by Proportional Counters.—

VI. The Beta Spectrum of Sulphur 35.

By A. L. COCKROFT and G. M. INSCH,
 Department of Natural Philosophy, Glasgow University*.

[Received July 18, 1949.]

SUMMARY.

The β -spectrum of sulphur 35 was analysed by means of a large proportional counter. The spectrum shows departure from the Fermi distribution for an allowed transition in the region of low energies. The spectrum has a maximum intensity at ~ 16.5 kev.

INTRODUCTION.

The β -spectrum of sulphur 35 was examined by means of the proportional counter with double anode described in Paper III. (Angus, Cockroft and Curran 1949). The counter was separated into two independent volumes by dividing the central wire into a long and short section joined together by a glass bead. The method of eliminating "end-effect" of the tube by taking the difference between the spectra as observed in the long and short parts of the tube has already been discussed. As the

* Communicated by the Authors.

end point of the spectrum is at 169 kev. the range of the particles of higher energy involves the use of a large pressurized counter. The counter used was almost identical with that described in Paper III., the pressure being at $5\frac{1}{2}$ atmospheres. The range of the most energetic β -particle was, therefore, of the order of one-fifth of the diameter of the counter.

Radioactive sulphur, S^{35} , was converted to hydrogen sulphide by heating a mixture of sulphur and paraffin wax. This procedure eliminated difficulties with the counter arising from the presence of the vapours of water or acids in the more usual "wet" methods.

The range covered in the present investigation extended down to ~ 7 kev. It is in the region of low energy that most spectrographic methods are least well established. With the aid of a large, high resolution magnetic spectrometer, Cook, Langer and Price (1948) investigated the β -spectrum of sulphur 35. They report an upper energy limit of 169.1 ± 0.5 kev. in agreement with previous accurate measurements. They observed that the experimental data for the various sources analysed agreed with Fermi theory for energies above about 76 kev., but that below this energy the experimental points rose above the theoretical straight line, and then fell rapidly for energies below 15 kev. The attenuation of the counter window probably accounts in part for the rapidity of this fall.

The value of 87 days for the half life of sulphur 35, together with the energy limit of 169 kev. seems to establish the β -transition as definitely allowed (Konopinski 1943). Moreover the complete failure to detect any γ -radiation associated with the disintegration seems to exclude all possibility that the spectrum is complex. It is particularly important, therefore, to establish with as much certainty as possible the nature of the disagreement with theory. The main purpose of this paper is to examine the low energy region and thereby to confirm that a discrepancy exists.

APPARATUS.

The counter has already been described in Paper III. The only significant difference was the division of the wire, so that the effective anodes were 45 cm. and 15 cm. long, giving a counter of effective length 30 cm. free from end effect. The counter was filled with a mixture of methane at a pressure of 30 cm. Hg., and argon at slightly more than 5 atmospheres. The counter was operated at voltages between 5000 and 6000 volts. As in the case of carbon the counter worked very satisfactorily under these conditions.

PREPARATION OF THE GAS.

The H_2S was prepared by heating a mixture of sulphur, paraffin wax and ignited asbestos. Preliminary control experiments showed that the H_2S reacted with the copper wall of the vessel forming the counter, but the use of a stainless-steel lining solved this problem. However, some of

the H_2S managed to penetrate behind the lining, and the counting rate decreased somewhat during the experiment. Care was taken to ensure that measurements which had to be compared as regards intensity were secured within a short time interval, thus eliminating any appreciable error that might arise from this effect. The partial pressure of the H_2S in the counter was about 1 mm. We had, previously, checked that this small admixture of stable H_2S with the argon and methane showed negligible electron affinity. This conclusion was based on the results of examination of the shape of the peak produced by homogeneous X-rays passed into the counter.

PULSE DISTRIBUTIONS.

The counting rates at the two ends of the counter were 19,000 and 5000 compared with background counting rates of 1400 and 400 counts per minute respectively.

To observe in detail the spectrum from about 5 kev. to 60 kev. the gas gain required was found to be $\sim 10^2$. The energy scale was calibrated by superimposing on the spectrum the fluorescence, $\text{K}\alpha$ X-rays of silver (22.1 kev.), obtained by reflecting a beam of white X-rays from a thin silver foil. Observations over the range 20 to 200 kev. were made with the same voltage on the counter, but with the amplifier gain reduced to $\frac{1}{4}$ of its previous value. As an additional check on the performance of the amplifier in the two conditions the fluorescence $\text{K}\beta$ X-rays of tungsten, energy 66.7 kev., were used as a calibrating radiation. These X-rays were fired directly into the counter through the copper wall (the $\text{K}\alpha$ group was strongly absorbed in traversing the copper). A third set of readings was taken for the long end alone, with a view to examining the spectrum in the vicinity of the end-point. The inset curve of fig. 1 gives the results. The resolution is not very satisfactory, mainly on account of the shape of the "background" spectrum, but a value of 168 kev. for the end-point seems most consistent with the experimental data.

These three sections of the spectrum separately examined were fitted together and the general result is shown in fig. 1. A Fermi Plot of the experimental curve was made (fig. 2). The theoretical Fermi distribution

$$S_{\text{TH}} = (W_0 - W)^2 (mc^2 + W)^2 / \{1 - \exp(-2\pi\alpha Z(W + mc^2)/p)\}$$

was used. Here W = kinetic energy of the β -particle, mc^2 = the rest energy, p = momentum, $\alpha = 1/137$ and $W_0 = 168$ kev. The Coulomb part of this expression is essentially non relativistic, but represents a good approximation even for energies of the order of the maximum energy of the spectrum.

DISCUSSION.

The results obtained here are in reasonable agreement with those of Cook, Langer and Price (1948). Both their curve and ours have a maximum at about the same energy value. Both sets of observations are in disagreement with Fermi theory for values below 75 kev. At higher

Fig. 1.

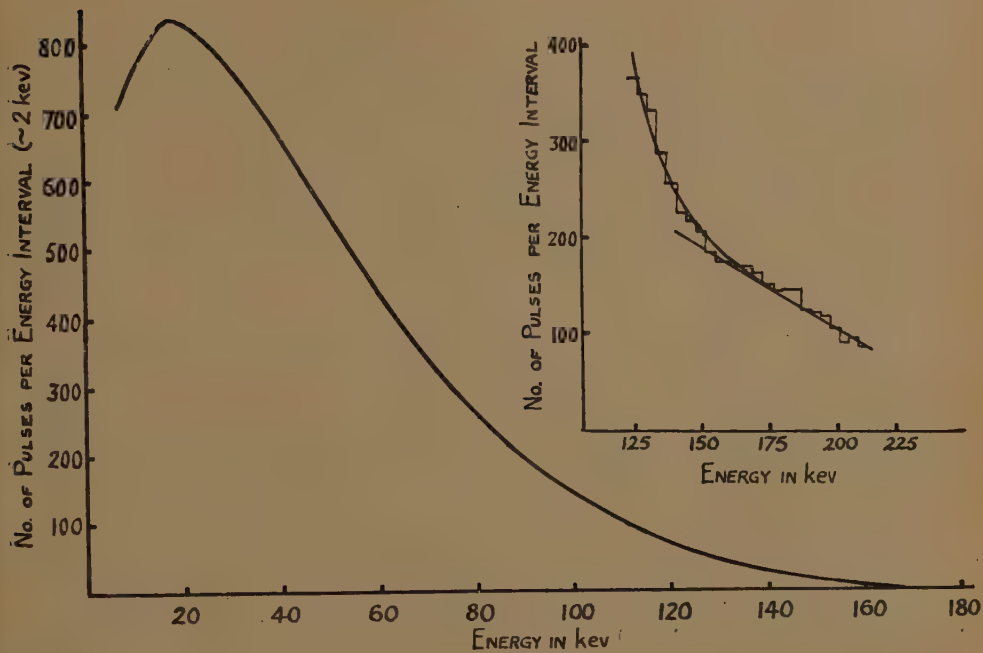
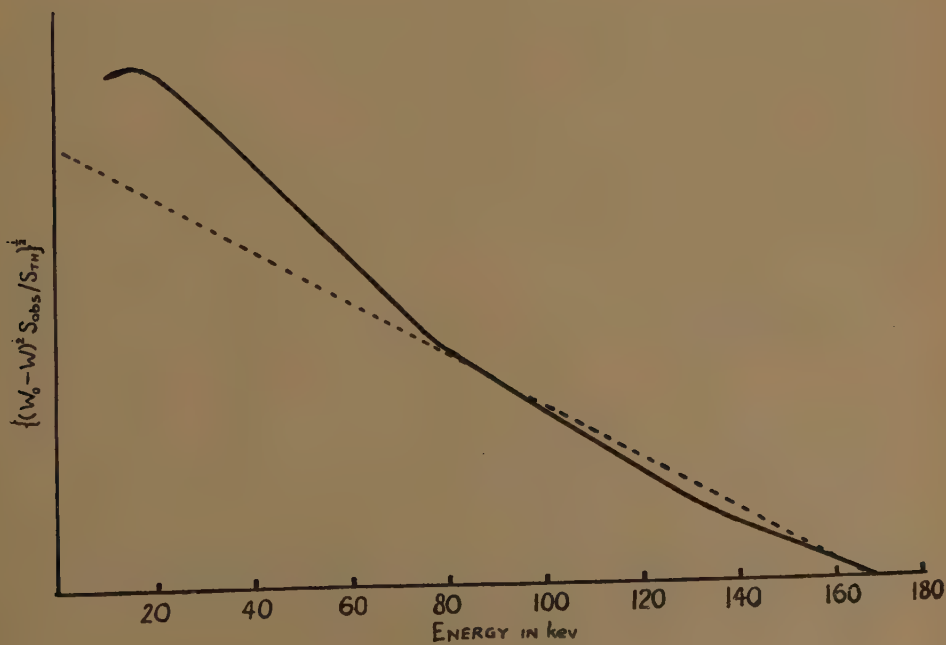


Fig. 2.



values of energy the agreement of our results with theory is less good, but we do not regard this difference as of any particular significance. Our results in this region are less satisfactory statistically, and the method of subtraction used here aggravates the difficulty. Moreover, the relatively large range of the high energy particles results in quite appreciable attenuation of measured intensity of the more energetic β -rays. An appreciable fraction of such particles strike the wall of the vessel, and do not expend all their energy in the gas. This "wall effect" gives rise to some displacement of the high energy pulses into the low energy region. Calculation shows that the experimentally observed deviation from Fermi distribution in the high energy region could be ascribed to such an effect. The resulting increase in intensity in the region of lower energy is relatively negligible, however, and the disagreement with Fermi here observed cannot be ascribed to this cause.

We conclude that Fermi theory is not in agreement with our experimental results at low energies, even though the β -transition of S^{35} is allowed. The observed departure from agreement is rather similar to that obtained in the work of Cook *et al.*, and the very great differences in the two techniques would appear to discount the possibility of a serious source of systematic error. We believe that the total error in measured intensities, due to all possible causes, cannot be greater than 10 per cent down to about 10 kev. Our results are not as accurate as those of Cook *et al.* in the upper half of the spectrum, but they are not inconsistent with their observations, which show agreement with Fermi in this region of the spectrum.

We wish to thank Professor P. I. Dee, F.R.S., for his help and advice throughout this work. We are grateful to Dr. S. C. Curran for useful discussions and assistance.

REFERENCES.

- ANGUS, J., COCKROFT, A. L., and CURRAN, S. C., 1949, *Phil. Mag.*, **40**, 522.
COOK, C. S., LANGER, L. M., and CLAY PRICE, H., 1948, *Phys. Rev.*, **74**, 548.
KONOPINSKI, E. J., 1943, *Rev. Mod. Phys.*, **15**, 209.

XCV. *Symmetry Changes in Barium Titanate at Low Temperatures and their Relation to its Ferroelectric Properties.*

By H. F. KAY and P. VOUSDEN,
H. H. Wills Physical Laboratory, University of Bristol*.

[Received July 26, 1949.]

SUMMARY.

The optical changes in BaTiO_3 are described, and it is shown that they can be completely explained if the crystal symmetry changes from tetragonal to orthorhombic at -5°C ., and then to rhombohedral at -90°C . These changes have been confirmed by X-ray investigations, and it is concluded that the three transitions are caused by the structure becoming successively spontaneously polarized along the [100] [110] and [111] cube directions. The relation of polarization to the cell structure is discussed and it is shown that the simple Lorentz equation is inapplicable if the titanium-oxygen interaction energy is large. An explanation of the three transitions is outlined on the basis of this interaction and some difficulties of electrical dipole cooperative effects in BaTiO_3 are discussed.

INTRODUCTION.

MEASUREMENTS of the temperature variation of dielectric constant of BaTiO_3 have revealed a principal peak at 120°C . and subsidiary peaks at -5°C . and -90°C . (Roberts 1947, Merz 1949). At 120°C . the crystal symmetry changes from cubic at higher temperature to tetragonal (Megaw 1947) and becomes spontaneously polarized along the tetrad axis (Hulm 1947). Harwood (1949) found discontinuities in the unit cell parameters at lower temperatures associated with the dielectric anomalies but it was thought that the symmetry remained tetragonal.

Changes in the birefringence also occur at these transitions (Kay 1948, Kay, Welland, Vousden 1949), and the present work is intended to give a more detailed description of these X-ray and optical changes and to discuss their relation with the ferroelectric properties of the material.

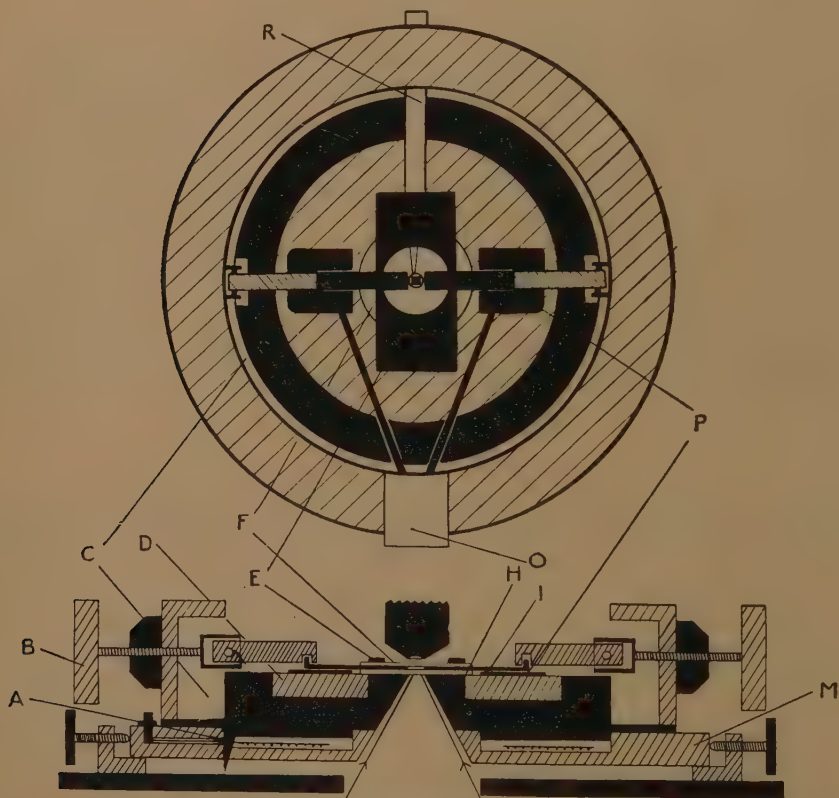
OPTICAL INVESTIGATION OF TRANSITIONS.

In order to examine BaTiO_3 crystals at temperatures down to -130°C . and also to apply electric fields across the crystals the apparatus drawn in fig. 1 was constructed. The crystal is placed between two glass coverslips (F) set on a cylindrical copper block (H), the electrodes (I) sliding in guides (P) fixed to an inset disc of insulating material (D).

* Communicated by the Authors.

Slight pressure is applied to the coverslips to ensure that the electrodes lie flat between them, the electrodes themselves being accurately ground so that fields can be applied to crystals as thin as 5×10^{-4} cms. Liquid oxygen can be pumped into the channel (C) containing cotton wool, and a heating coil (A) is contained in the asbestos disk (M) beneath the block. This arrangement is fitted on the stage of a polarizing microscope, an objective magnification of $\times 40$ being used. Crystals can rapidly be cooled to -130°C ., a thermocouple (R) recording the temperature.

Fig. 1.



Apparatus used for the application of fields to crystals at different temperatures. Further key : (B) Control for movement of electrodes.

(E) Metal plate to secure coverslips.

(O) Leads and plug for high voltage supply (± 2000 volts).

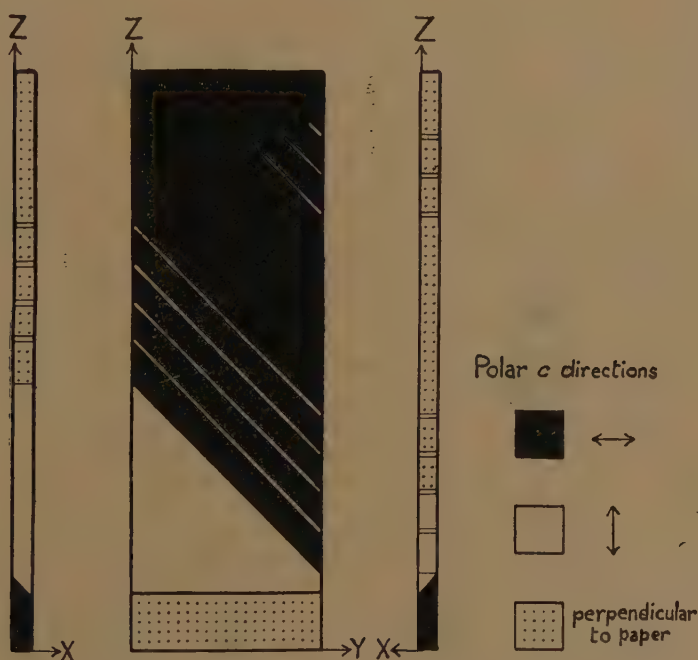
Full shading : brass or copper.

Line shading : electrical or thermal insulating material.

The crystals of BaTiO_3 investigated were of plate form with $\{100\}$ faces developed, the areas of the plates being $\sim 10^{-3}$ cms.² and their thickness $\sim 10^{-3}$ cms. This order of thickness was essential for the clear interpretation and measurement of the changes in birefringence.

At room temperature crystals of BaTiO_3 are tetragonal pseudo-cubic and are usually twinned across the (101) or (011) planes. In thin crystals the individual boundaries may be seen and the orientation of the components completely determined, the c axes lying nearly parallel to two or three of the cube axes of the macrocrystal, or in some cases the crystals are completely untwinned. Fig. 2 shows the arrangement actually observed in a crystal having dimensions $X=0.01$, $Y=0.1$, $Z=0.33$ mm. containing fourteen twin components twinned in three directions. Of interest are the very thin twin lamellae in the (011) plane whose boundaries are separated by $\sim 2 \times 10^{-4}$ cm., this value being approximately constant and independent of the crystal size. Such thin twins are frequently observed and may extend across the crystal or terminate within it by oblique junction.

Fig. 2.

Typical twinning in BaTiO_3 at room temperature.

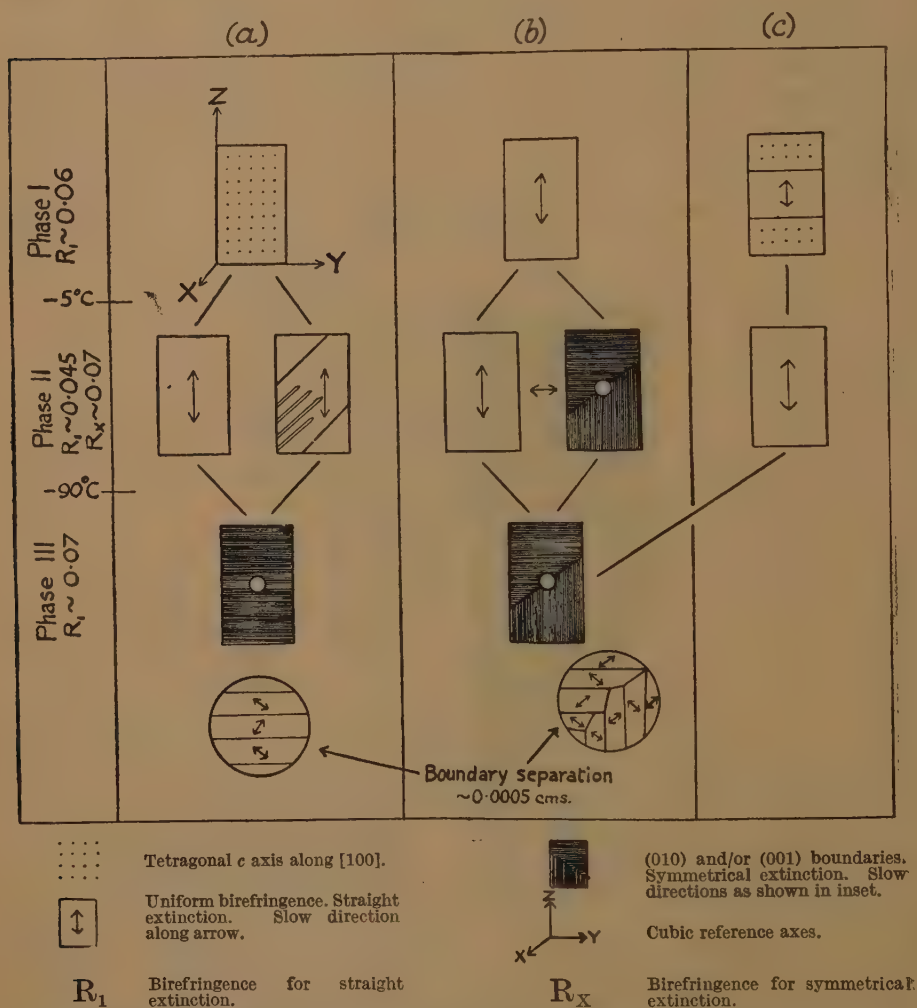
At -5°C . and -90°C . crystallographic changes occur which are rather complex and which depend on the initial orientation of the c axis in the crystal plates. The changes of twinning, birefringence and extinction direction for the three simplest types of crystal are shown in fig. 3 where the plates are viewed normal to their plane.

Below -5°C . (phase II) the extinction may be straight or symmetrical. In the latter case fine-scale twinning always occurs, and in the former, twinning where present is apparently similar to that at room temperature

often showing the very thin lamellae. Below -90°C . the extinction is always symmetrical and the crystal still highly twinned, no further change occurring down to -130°C .

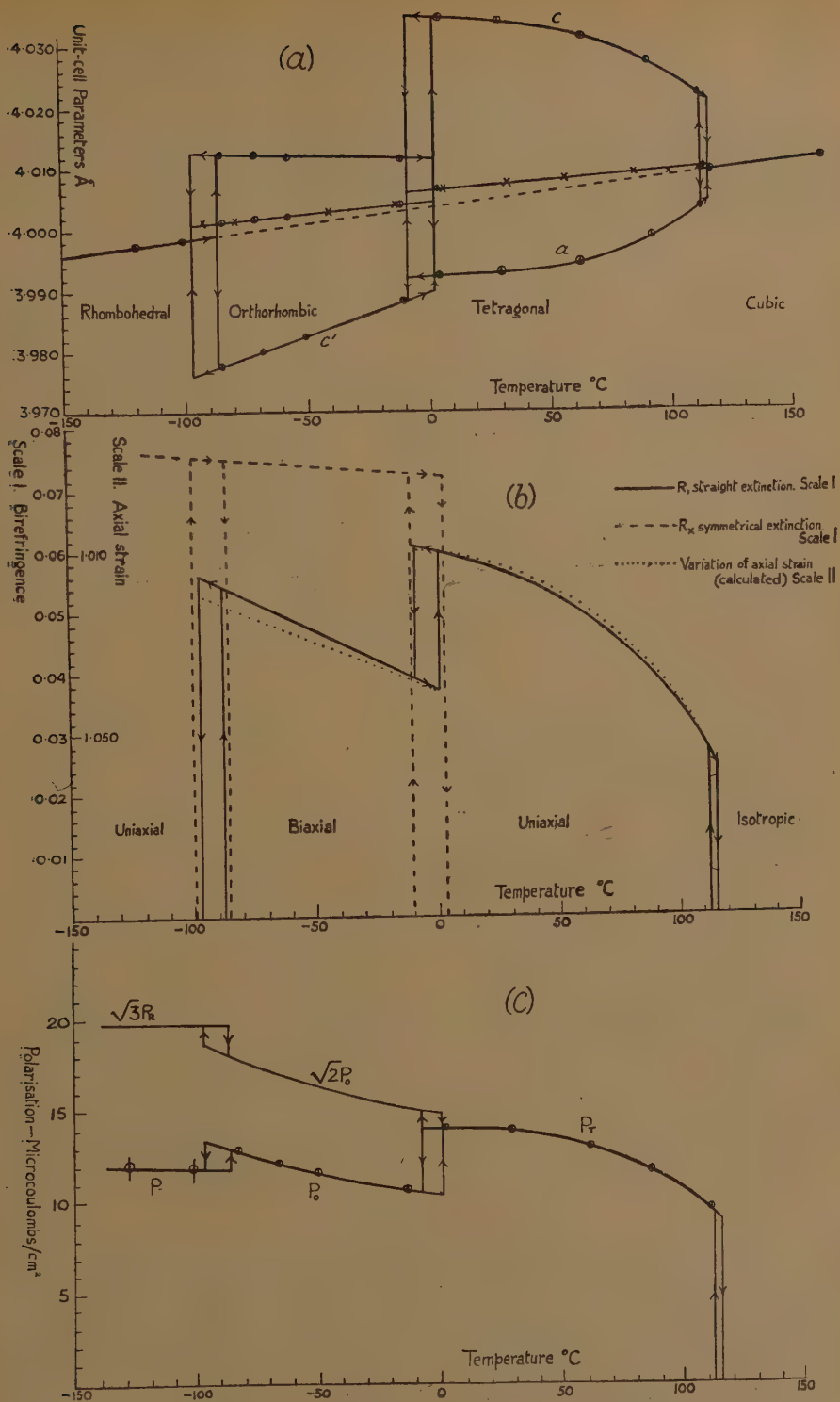
The relation between the orientation of the tetragonal c axis at room temperature and the extinction in phase II is characteristic. Thus with

Fig. 3.



Appearance of simply twinned BaTiO_3 crystals with polarized light in transmission position at temperatures corresponding to the three main phases. crystals of type (a) and (c) the extinction is straight and any room-temperature twin boundaries completely disappear. In type (b) the extinction is usually symmetrical and the transition very sharp, but if the extinction remains straight the birefringence is reduced and phase I and II coexist over a small temperature range. At both transitions there is

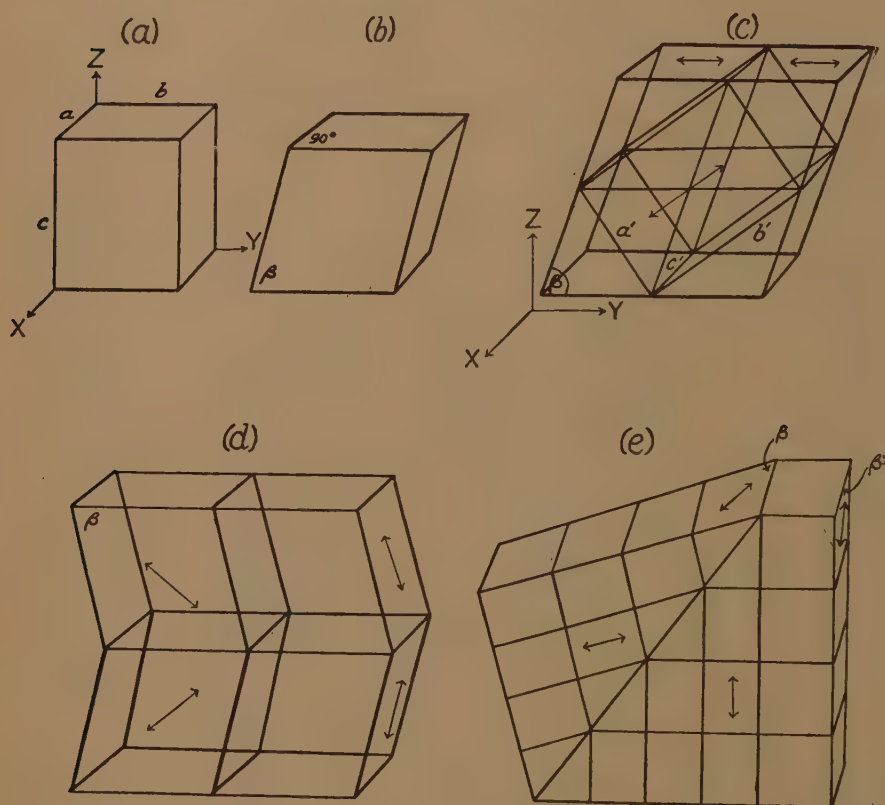
Fig. 4.



- (a). Variation of cell parameters with temperature in BaTiO_3 crystals. The thick line corresponds to the cube root of the volume of the cell and the crosses its calculated value.
- (b). Variation of birefringence (R) with temperature.
- (c). Variation of spontaneous polarization with temperature.

considerable temperature hysteresis ($\sim 9^\circ \text{C.}$), this being greatest for crystals transforming with symmetrical extinction which, together with their sharp transition, suggests maximum resistance to such a change. All types of crystal return to their original room temperature form, sometimes with simplified twinning.

Fig. 5.



Relationship between cell shape and orthorhombic twinning of BaTiO_3 .

X , Y , Z are the external cube axes; (a) tetragonal cell, (b) the sheared primitive cell and (c) the orthorhombic cell. (d) and (e) show the possible twinning of the orthorhombic cell with arrows giving the extinction direction.

The variation of birefringence (R) with temperature for both cases of extinction was measured by compensation methods and the results are plotted in fig. 4(b). The probable error of the results is ± 5 per cent, except below -90°C. when the fine-scale twinning considerably reduces the accuracy.

The observed changes can be unequivocally explained if at -5°C. the unit-cell distorts from tetragonal pseudo-cubic (fig. 5(a)) by a small shear in the (100) or (010) plane. The two axes in the shear plane become

equal, with magnitude between the original a and c parameters and the third axis remains near its room temperature value. The new cell is orthorhombic with its axes a' , b' , c' in the directions $[101]$ $[011]$ and $[100]$ (or their equivalent) relative to the crystal axes X , Y , Z (figs. 5(b) and (c)). The extinction of this cell is straight if viewed along the cube axes Y or Z and symmetrical if viewed along the X -axis. Thus we can explain the two optical orientations of phase II and the rules governing their relative orientation to the tetragonal cell.

The methods by which the orthorhombic cell may twin are shown in figs. 5(d) and (e). In 5(d) the twin planes are (110) and may be present in two directions inclined at $90 \pm (\text{angle of shear } (\delta))$ as in fig. 3(b). In fig. 5(e) the twin plane is (111) and the twinning similar to that of the tetragonal form, but because of the shear, the planes of the two twin components drawn in the plane of the paper are actually inclined to it at a small angle. This twinning may occur in the transition of crystals of type shown in fig. 3(a).

From the observed values of the birefringence for straight extinction the axial strain may be calculated since the relation between strain and birefringence is known in the tetragonal form. The shearing of the cell does not alter this birefringence (R), as the ellipse representing the shear strain has its plane containing the direction of the incident light, and its axes equally inclined to it. Thus the spacings along the cube axes just below -5°C. are found to be in the ratio $1.0065 : 1.0 : 1.0065$ compared with their room temperature values $1.000 : 1.000 : 1.010$. The angle of shear cannot be found from measurements of R but a rough value ~ 15 minutes was obtained from measurements of the angle between twin boundaries. From the small variation of birefringence with temperature (fig. 4) it is seen that δ is nearly constant.

Below -90°C. the extinction is always symmetrical. It was therefore concluded that the crystal symmetry is rhombohedral as this is the only pseudo-cubic form which would give twinning and extinction independent of the cube axis of viewing. The cell is now stretched (or compressed) along a cube diagonal, all cube faces being slightly sheared, and the only possible twinning planes are parallel to these faces, the appearance of a crystal being similar to that of the orthorhombic form viewed with symmetrical extinction.

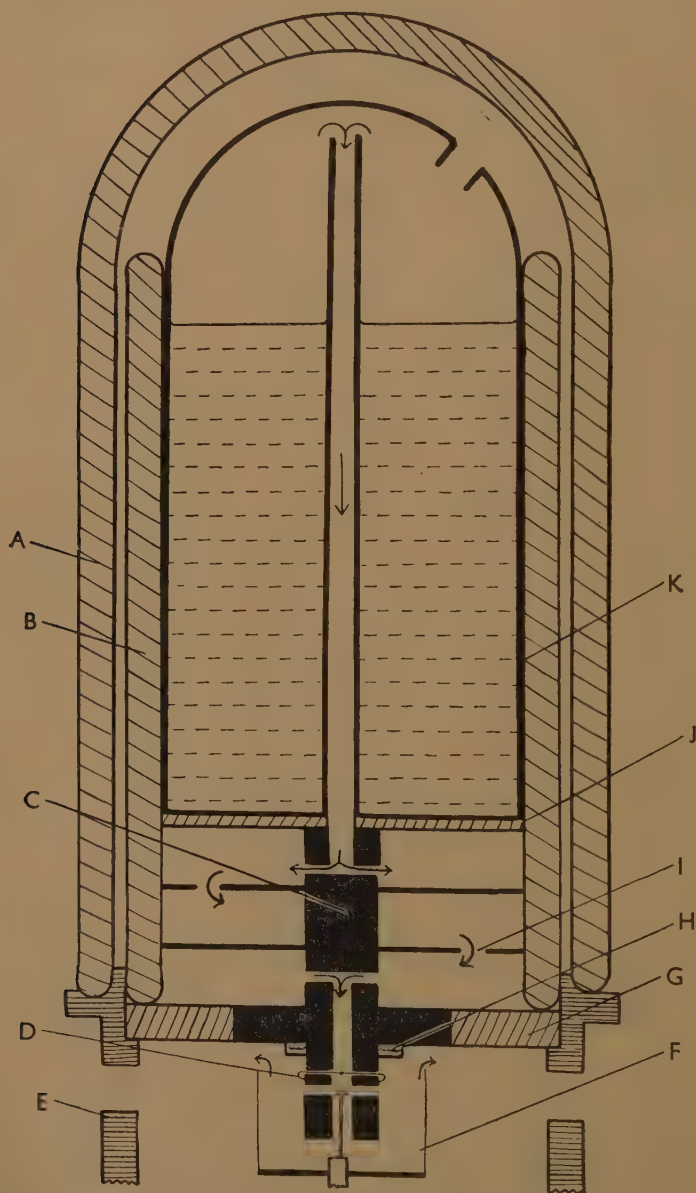
* * * * *

X-RAY CONFIRMATION OF SYMMETRY CHANGE.

The cooling apparatus (fig. 6) was in some respects similar to the type already described by Ubbelohde and Woodward (1947), but of very simple design and suitable for use with a Unicam goniometer. A drilled copper rod (C) and cellophane cup (F) surrounds the crystal, slots being cut for the incident and emergent X-ray beams. This contains an inset thermocouple (D) wound to reduce conduction along its length, and a heating coil (H) for small temperature changes ($+25^\circ \text{C.}$). The rod is held in

position by a Tufnol cylinder (E) and the cork (G) which fits exactly into the recess normally occupied by the cylindrical film holder. The crystal

Fig. 6.



Crystal cooler for use in conjunction with standard X-ray goniometer.

is cooled to -50°C . by conduction through the rod with CO_2 in the Dewar cylinder (B), lower temperatures being reached with liquid oxygen

contained in the copper can (K) fitted inside. In this case the crystal is further cooled by the flow of cold gas down the central tube, the interchange of heat occurring through the baffles (I) before the gas passes over the crystal. This indirect flow was found necessary to eliminate difficulties of temperature variation and icing up of the crystal when the tank was replenished.

Large temperature changes are effected by the insertion of thin discs of poorer conductivity (J) between the can and rod and an outer inverted dewar (A) covers the whole system. The temperature of the crystal can be maintained without attention to $\pm 1^\circ \text{C.}$ at -50°C. for 6 hours using CO_2 or at -100°C. for 2 hours with liquid oxygen, and refilling with coolant after these periods can be effected without appreciable temperature change.

The collimator was fixed at a distance of 10 cm. from the crystal together with a back reflection plate. A multiple exposure rotating film holder was fitted at 90° to the incident beam to record the 330 group of reflections ($\theta \sim 55^\circ$), and the whole apparatus allowed the crystal to be observed with polarized light with a magnification of $\times 14$.

To confirm the symmetry changes deduced from the optical investigation, it is only necessary in such a pseudo-cubic material as BaTiO_3 to examine reflections from planes approximately normal to the cube axes and face diagonals of the unit cell. A single crystal technique is essential, for with powder photographs the line multiplicity is too great, *e.g.* the group $h^2+k^2+l^2=26$ contains 20 lines when the crystal symmetry is orthorhombic. This line multiplicity together with the close twinning of the structure will cause considerable apparent line broadening and this is probably one of the main causes of the increased diffuseness of powder reflections obtained by previous workers. Only the essential photographs and orientations of the crystal are described here for the sake of brevity.

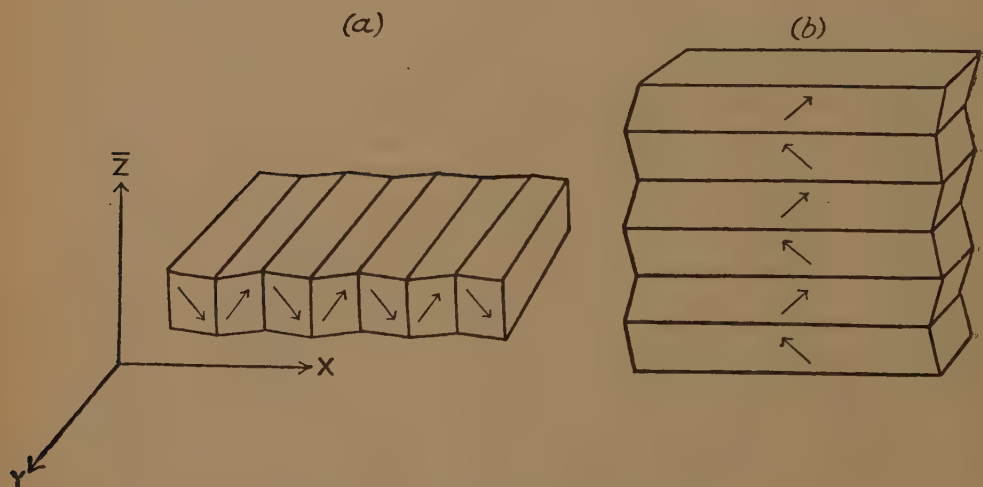
To obtain the best possible resolution it was decided first to investigate crystals having sharp transitions, uniform and complete extinction, and which appeared untwinned below -5°C. This permits the use of a crystal with some degree of twinning at room temperature which in fact helps to determine the crystal orientation below -5°C. The dimensions of the crystal used were $0.2 \times 0.2 \times 0.03 \text{ mm.}$ and was similar to that shown in fig. 3(c). The geometrical broadening of reflections was reduced as much as possible by employing large crystals to film distances ($\sim 10 \text{ cm.}$) and small collimator aperture ($\sim 0.15 \text{ mm.}$ diameter) together with the small crystal. Exposure times were still quite short ($\sim 7 \text{ m.a.h.}$) this being partly achieved by the use of small oscillations of the crystals ($1^\circ - 5^\circ$).

To detect and measure the shear in the ac plane below -5°C. the crystal was mounted with this plane normal to the rotation axis (*i.e.* the Y axis *cf.* fig. 3(c) and fig. 7(a)), and oscillation photographs were taken of the cubic groups of reflections 501 and 300 at different temperatures on the same film.

The results show that at -7°C . the structure is sheared by 8.4 ± 1.5 minutes and the original a and c parameters become equal and less than the value of c at room temperature. The parameter changes were discontinuous and the crystal was found to be closely twinned as in fig. 7(a). This twinning is undetectable by optical examination as there is no change in refractive index across the twin boundaries when viewed normal to the plate.

To obtain the new spacing perpendicular to the shear plane a crystal showing symmetrical extinction with twin boundaries in only one direction was mounted with these boundaries horizontal (fig. 7(b)). Reflections from planes nearly parallel to the plate enabled the value of c' to be determined. From planes nearly normal to the plate the longer spacing was obtained and the reflections showed vertical splitting resulting from

Fig. 7.



Small arrows indicate polar direction.

Orientation and twinning of crystals used in X-ray investigation, illustrating the results of the application of fields to crystals below -5°C .

the twinning, from which the twin angle and hence the angle of shear could be calculated. The photographs showed that the twinning was not always homogeneous across the thickness of the plate ($\sim 10^{-2}\text{ cm.}$) as had been supposed but the value of δ obtained $9' \pm 0.5'$ at -30°C . agreed with that from spacing determinations within experimental error. The slight inhomogeneity of twinning below -5°C . is explicable for such comparatively thick crystals, as the sharpness of the transition makes an exact fit between all twin components very unlikely unless the crystal thickness is comparable with the twin boundary separation.

Below -90°C . it was found that all reflections could be indexed on the assumption that the symmetry is rhombohedral, all cube spacings being equal and the cube faces sheared through a small angle γ . The changes

in all parameters were discontinuous, the new spacing parallel to the cube faces lying between the original a and c parameters. The shear in these faces may be calculated from the two appropriate spacings or from the vertical splitting of reflections due to twinning. The average value obtained was $7.5 \pm 0.5'$, all methods agreeing within experimental error.

A complete graph of the temperature variation of spacings parallel to the cube axes is drawn in fig. 4(a) the accuracy being 1 part in 10^4 . In fig. 4(b) the variation of the axial ratio, or more strictly the relative axial strain, is shown together with the values obtained from the birefringent measurements, the two curves agree within experimental error. The variation of shear angle (δ) has only been accurately measured down to -40°C. when its value is $8.75' \mp 0.15'$ compared with $8.4 \mp 0.15'$ at -7°C. This small variation also agrees with the optical measurements. The change of δ at -90°C. is $1' \mp 0.5'$ and cannot be measured optically because of the close twinning.

The results of the X-ray investigation were in complete conformity with the optical results and explain the powder photographs of previous workers. In particular the increase in intensity of the $00l$ compared with the $h00$ reflections as the temperature is lowered (Megaw 1947) has a simple explanation in that with the orthorhombic symmetry there are twice as many contributing planes for the longer spacing than for the shorter, thus reversing the relative intensities of reflections compared with room temperature. The decrease of spot sharpness with temperature is quite small, reflections differing in spacing by 1 part in 2000 being resolvable at -100°C. , and it seems, therefore, that the diffuseness in powder photographs is largely due to line multiplicity and strain inhomogeneity resulting from irregular twinning. All three transitions are of first-order type accompanied by a volume decrease of $0.1 \mp 0.04 \text{ \AA}^3$ at -5°C. and -90°C.

APPLICATION OF ELECTRIC FIELDS TO CRYSTALS.

At temperatures between -5°C. and 120°C. an electric field applied along a cube axis will orientate the tetragonal axis of twins parallel to the field, saturation occurring near the breakdown field strength of $\sim 2 \times 10^4$ volts/cm. whereas a field applied along the $[110]$ axis causes no observable orientation. Both these results are readily explained if the material is spontaneously polarized along the tetragonal axis and the saturation field of the same order as the breakdown value.

At temperatures between -90°C. and -5°C. fields applied along $[100]$ cause no change if the crystal has symmetrical extinction, but if the extinction is straight the slow (original c) direction is aligned parallel to the field, the process being accompanied by twinning similar to the tetragonal case. If, however, a field is applied along $[110]$ the extinction is changed from straight to symmetrical, the crystal becoming largely untwinned and having the slow direction parallel to the field. The twinning is reduced with increasing field and only one direction of the applied field will effect this change.

These results can be explained if the direction of the polarization is now along the a' or b' axis of the orthorhombic cell. Thus in fig. 7(a) is shown a twinned crystal with straight extinction, the two directions of the polar axis being $[101]$ and $[10\bar{1}]$ referred to the external axes. Applying a field along $[010]$ would tend to orientate the polar direction in the $[011]$ and $[01\bar{1}]$ directions by a mechanism similar to that at room temperature, the extinction remaining straight, and the slow direction being parallel to the field. The application of a field along $[110]$ would orientate the polar axis completely in this direction, the shear plane now being XY, but not if applied along $[\bar{1}\bar{1}0]$ as then the coercive field is much higher, for the polarity must first be reversed before alignment can occur.

The slow direction is again parallel to the field and as this is still the direction of extensive strain we may further conclude that the polar axis is parallel to b' and not the a' axis of the orthorhombic cell (fig. 4(c)).

At 120°C . the change to tetragonal symmetry is due to the spontaneous strain accompanying the spontaneous polarization, and similarly below -5°C . the polarization along the $[110]$ cube axis will distort the cell to the orthorhombic form, the forces being of an electrostrictive nature and the strain proportional to the square of the polarization. Below -90°C . the rhombohedral symmetry may now be explained if the polar direction is $[111]$, the polarization now extending (or compressing) the cube diagonal of the unit cell. The sign of the strain could not be determined from the X-ray investigation because of the twinning, but since in the orthorhombic form the polarization is along the extended b' axis in the sheared plane, below -90°C . the other two cube faces will be similarly sheared, and therefore the cube diagonal extended by the polarization. It is unfortunately very difficult to confirm this direction of polarity as no crystals with $\{111\}$ faces were available. Fields applied along $[100]$ cause no orientation before breakdown occurs, a result consistent with a $[111]$ polar axis, as the twinning from a simple crystal at room temperature allows four different polar directions equally inclined to the field below -90°C .

This change of polar direction may be regarded in another more significant way, namely, that the crystal structure is successively polarized along the cube axes at the three transitions giving the observed resultant polarities. Thus in cooling a crystal below -5°C . the polarization in the plane of the plate, originally parallel to the X-axis of fig. 7(a) is only slightly altered in magnitude, and together with this there now appears an equal and additional polarization perpendicular to the crystal plate which is opposed in adjacent twins. Below -90°C . the same type of change will occur resulting in additional twinning and the resultant $[111]$ polarity.

These observations throw new light on the domain structure of ferro-electrics. In Rochelle salt the polarization between the Curie points causes a shearing of the orthogonal unit cell in the plane normal to the polarization; adjacent twins have opposite polarities and their greatest

length is along the polar axis. It is natural to suppose that this domain shape is caused by the fact that the "de-electrifying field" is least for this form, or alternatively that thereby neutralization of polarity by the collection of surface charges can be most easily effected. If this were the case we should also expect barium titanate to consist of a similar antiparallel array of domains in the tetragonal form. However, from the one way effect of the field below -5°C . we may conclude that the polarization in the plane of the plate is unidirectional and hence that the crystal is unipolar at room temperature. This effect has been found for all the optically single crystals so far examined, and a detailed optical examination has not as yet shown any evidence of strain at boundaries which might be expected from an antiparallel domain structure.

Though this evidence is not conclusive, it strongly suggests that the observed domain structure in BaTiO_3 and Rochelle salt is due to other than electrical forces. An alternative explanation is that the elastic stresses caused by one portion of a crystal transforming cause the adjacent portion to transform as a related twin, the opposed polarity being a result of this twinning. This is equivalent to the fact that twinning in pseudo-cubic materials is such as to mimic the higher symmetry, thus causing least change in external crystal shape. This is the normal type of twinning occurring in the tetragonal phase of normal BaTiO_3 and this does not result in a state of lower electrical energy. A detailed description of the twinning in this phase and in particular of a square network twin arrangement is to be published later.

DISCUSSION OF RELATION OF SYMMETRY CHANGES TO THE FERROELECTRIC PROPERTIES.

Theories advanced to explain the properties of BaTiO_3 have so far centred round the abnormally large volume available for the titanium ion, a point discussed in some detail in the dielectric conference (*Trans. Far. Soc.* 1946) and the evidence for which is strengthened by a study of related structures, *e.g.* CaTiO_3 and Ba, SrTiO_3 . Rushman and Scrivens suggested that this might cause the ion to be displaced from the centrosymmetrical position giving the cell a permanent dipole moment and an order-disorder transition. At that time the very high value of the dipole moment per unit cell was unknown ($\sim 4 \times 10^{-18}$ e.s.u.), and such ion shifts were thought unnecessary to explain the dielectric behaviour at 120°C ., but it is now clear that to explain this value shifts of the order of 0.1 \AA must occur.

A permanent titanium ion displacement has formed the basis of a theory (Mason and Matthias 1948) put forward to explain the spontaneous polarization in which a six-position potential well was assumed, the minima being displaced 0.15 \AA from the cell centre toward six oxygens. A Langevin-Weiss approach was used to calculate the statistics of the model, the Lorentz internal field providing the cooperative field. Though calculations of the various constants would seem to agree satisfactorily with experimental values (see, however, Merz 1949, Roberts 1949), the

theory is open to some objections of which the most relevant to the immediate discussion is that the model only gives one transition point. This is because it is assumed the only cooperative effect is due to the interaction of the six-position Ti dipole and the Lorentz internal field, so that the orientation of the dipole by the field at 120°C. results in the lowest energy state being attained. Consequently no other transitions are likely on physical grounds, and the objection applies to any model containing only Ti dipoles, unless we gratuitously introduce a potential well system which would give the observed changes in dipole direction.

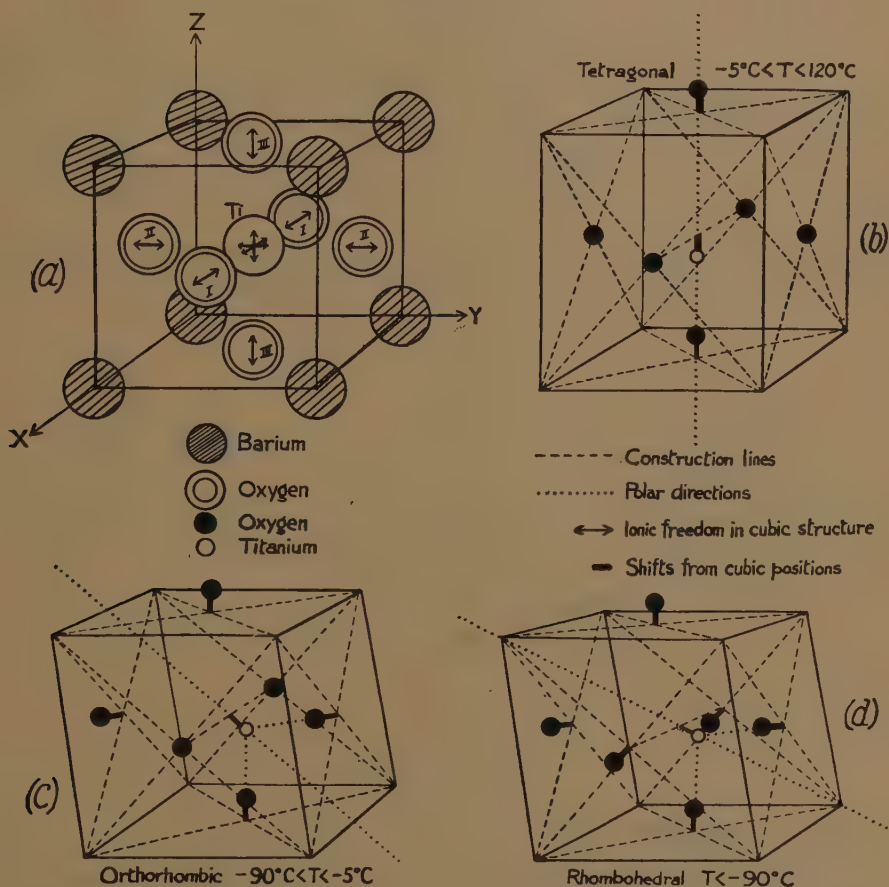
An alternative model is now discussed which is supported more strongly by the structural data and gives rise to the three transitions observed, but it must be remembered that in view of our lack of detailed knowledge of charge distribution in the crystal any such model must be a very rough approximation. On the other hand, thermo-dynamical coordination can relate the dielectric constants and strain to the polarization, so that a model need only account for the source of polarization and its temperature dependence.

It is here assumed that the determining factors in the crystal structure which cause it to be ferroelectric below 120°C. are the shape of the potential wells in which both the highly charged Ti ion and highly polarizable O ion move. The ion polarizability is inversely proportional to the absolute temperature for ions in simple types of discrete position or flat potential wells, and we may on general grounds expect them to give a transition for these cases when the interaction energy of neighbouring ions is of the order of kT . This interaction is proportional to μF where F is the local field and μ the effective dipole moment of the ion which is constant with temperature and varies slightly by a numerical constant depending on the exact type of well considered. On the other hand, if the potential well is parabolic and the ionic vibrations harmonic, then the interaction energy decreases linearly with temperature and no transition will occur.

Approximate calculations of the shape of the Ti and O potential wells from the atomic binding forces show that for both ions the well is rather flat and that there may be a two-position well at the oxygen position displacing it towards the titanium (A. F. Devonshire). This conclusion is also supported by the cell parameters, for as well as the Ti-O distance being too large the Ba-O distance is too small, a situation which could be cured by a displacement of the oxygen towards the Ti giving it two equilibrium positions. Since the oxygen ion is easily distorted, calculations from Goldsmidt radii are of doubtful value in obtaining a value for this shift, but assuming the true Ba-O distance to be equal to their valency-corrected Goldsmidt sum, then from the observed cell parameters a simple calculation shows the oxygen ion could move $\sim 0.2\text{ \AA.}$ This would bring the Ti-O ions closer than their Goldsmidt radii allow and might be caused by a covalent bond between them, a suggestion already made to account for the dielectric behaviour of other titanium compounds, *e. g.* TiO_2 (*cf.* Mott and Gurney).

To sum up we may say the evidence supports an abnormal degree of freedom of both the ions as in fig. 8(a), and even above 120°C . there may be a permanent displacement of the oxygen from the face-centre with a small potential barrier separating the two positions. From the discussion on the influence of potential well shape on ion polarizability it follows that

Fig. 8.



Changes in symmetry and proposed ion shifts in the primitive cell of BaTiO_3 .

to a first approximation the ion contribution to the polarization can be expressed by assigning dipole moments μ_{Ti} , μ_{O} to the Ti and O ions. The change of crystal symmetry to tetragonal and the subsequent polarizing field will modify these values of μ below 120°C ., but since the increase in cell length is only $\sim 0.015 \text{ \AA}$ at the transition this effect should be small and can be adequately treated on a phenomenological basis.

In any statistical treatment of the behaviour of the dipole system outlined it can be shown at once that the simple Lorentz equation is invalid, for even though the crystal symmetry is cubic above 120°C . the dipole array has tetragonal symmetry. If in fig. 8(a) we consider the

polarizability of the unit cell along an arbitrary cube axis X then the contribution of the oxygen ions II and III are very much less than that of I. This will give rise to a short range dipole interaction energy which is of the same order of magnitude as the Lorentz polarization energy. Its value may be simply calculated remembering that the mutual potential energies of parallel dipoles $\mu_1 \mu_2$ at a distance r with axes inclined to r at an angle θ is

$$\frac{\mu_1 \mu_2}{r^3} (3 \cos^2 \theta - 1).$$

For arrays of point dipoles having cubic symmetry this expression summed over all dipoles within a sphere of arbitrary radius is zero. Thus in BaTiO_3 the $\mu_{\text{Ti}} - \mu_{\text{Ti}}$ and $\mu_0 - \mu_0$ are based on a cubic array and the interaction is zero whereas the $\mu_{\text{Ti}} - \mu_0$ gives rise to a short range aligning field.

Selecting a Ti dipole and summing its interaction $E_{\text{Ti}-0}$ with the surrounding oxygen the result for the first two shells is

$$E_{\text{TiO}} = (16 - 1.6 + \dots) \frac{\mu_{\text{Ti}} \mu_0}{d^3} \\ \doteq 16 \frac{\mu_{\text{Ti}} \mu_0}{d^3}.$$

where d is the cell parameter.

The contribution falls off very rapidly as the distance from the external dipoles increases and to a very good approximation the interaction energy may be taken as that between nearest neighbours. The Lorentz polarization energy (E_L) per unit cell on this model is

$$E_L = \frac{2}{3} \pi d^3 P^2 \quad \text{and} \quad P = \frac{\mu_{\text{Ti}} + \mu_0}{d_3}.$$

$$\therefore E_L = \frac{2}{3} \pi (\mu_{\text{Ti}} + \mu_0)^2 / d^3.$$

Hence

$$\frac{E_{\text{TiO}}}{E_L} = \frac{24}{\pi} \frac{\mu_{\text{Ti}} \mu_0}{(\mu_{\text{Ti}} + \mu_0)^2}.$$

This ratio depends on the value of $\mu_{\text{Ti}}/\mu_0 = 2\delta_{\text{Ti}}/\delta_0$ where δ_{Ti} , δ_0 are the atomic shifts. The only published evidence of their magnitudes is that $\delta_{\text{Ti}} \sim 0.15 \text{ \AA}$ (Danielson 1948), and also that $\delta_{\text{Ti}} < 0.05 \text{ \AA}$ with $\delta_0 < 0.13 \text{ \AA}$ (Kay, Wellard, Vousden 1949).

In view of these conflicting results, no definite conclusions may be drawn, but taken in conjunction with the value of the spontaneous polarization $[=(\mu_{\text{Ti}} + \mu_0)/d^3]$ seem to point to the oxygen shift being larger than the titanium and of order of magnitude 0.15 \AA .

We may expect then $\mu_{\text{Ti}}/\mu_0 \sim 1$ and thus $E_{\text{TiO}}/E_L \sim 2$, so that the short range interaction is the predominating term and even if $\mu_{\text{Ti}}/\mu_0 = \frac{1}{5}$ or 5 as extreme cases we have still $E_{\text{TiO}}/E_L \sim 1$.

On this basis a qualitative explanation of the three transitions may be made. When the interaction is of the order of kT the crystal becomes spontaneously polarized along an arbitrary cube axis, the resultant

internal field straining the crystal along this direction (fig. 8(b)). The reduction in free energy due to the strain lowers the Curie point in the other two axial directions to -5°C . when the structure becomes polarized along an additional cube axis, giving the resultant $[110]$ polar direction and orthorhombic symmetry (fig. 8(c)). This strain again lowers the third Curie point to -90°C . when the resultant polarity is along $[111]$ (fig. 8(d)). The lower transition temperatures must further be lowered since in the tetragonal phase the effective dipole moment of the Ti ion perpendicular to the tetrad axis will be reduced, and a similar effect will occur at -5°C . The magnitude of this reduction will depend on the value of δ_{Ti} along the tetrad axis, as the shift will alter the shape of its potential well in the perpendicular directions so as to reduce the short range interaction.

Since we have no detailed evidence of atomic shifts the quantitative statistical theory is not developed further in this paper, but assuming $\mu_{\text{Ti}} = \mu_0 \sim 10^{-18}$ e.s.u. cm., which gives $\delta_{\text{Ti}} = 0.05 \text{ \AA}$ and $\delta_0 = 0.1 \text{ \AA}$ and $P = 10$ microcoulombs/cm.², then $E_{\text{TiO}}/kT \sim 4$ which is the right order of magnitude.

* * * * *

We may, however, draw conclusions supporting this model from an analysis of the electrostrictive effect, the treatment differs slightly from Devonshire (1949) in the analysis of the volume change.

Expanding the free energy (A) as a function of strain and polarization and taking our zero at T_0 defined by the Curie Weiss law, $\chi = K/T - T_0$, where χ is the susceptibility, we have, including second-order terms,

$$\begin{aligned} A = & \frac{1}{2}c_{11}(x_x^2 + y_y^2 + z_z^2) + c_{12}(x_x y_y + y_y z_z + z_z x_x) + \frac{1}{2}c_{44}(x_y^2 + y_z^2 + z_x^2) \\ & + \chi^{-1}(P_x^2 + P_y^2 + P_z^2) + \frac{1}{4}b(P_x^4 + P_y^4 + P_z^4) + B_{11}(P_x^2 x_x + P_y^2 y_y + P_z^2 z_z) \\ & + B_{12}[P_z^2(x_x + y_y) + P_y^2(z_z + x_x) + P_x^2(y_y + z_z)] + \frac{C_{11}}{2}(P_x^2 x_x^2 + P_y^2 y_y^2 + P_z^2 z_z^2) \\ & + \frac{C_{12}}{2}[P_z^2(x_x^2 + y_y^2) + P_y^2(z_z^2 + x_x^2) + P_x^2(y_y^2 + z_z^2)] \\ & + D(P_x P_y x_y + P_y P_z y_z + P_z P_x z_x) + \alpha(c_{11} + 2c_{12})(T_0 - T)(x_x + y_y + z_z), \end{aligned}$$

where P_x, P_y, P_z are the permanent polarizations, x_x, x_y, x_z , etc., the elastic strains, c_{11}, c_{12}, c_{44} the elastic coefficients, B_{11}, C_{11} etc., the coefficients of polarizability and α the thermal expansion coefficient.

Minimizing with respect to the strains and solving for x_x, y_y, z_z we find for the rhombohedral phase in which $x_x = y_y = z_z$ and $P_x = P_y = P_z = P_R$

$$\begin{aligned} \alpha(T_0 - T)(c_{11} + 2c_{12}) + (c_{11} + 2c_{12})_R x_x + P_R^2(C_{11} + 2C_{12})_R x_x \\ + (B_{11} + 2B_{12})P_R^2 = 0. \end{aligned}$$

Now, from fig. 3(a), we see that

$${}_R x_x = \alpha(T - T_0).$$

Hence

$$(C_{11} + 2C_{12})\alpha(T - T_0) + (B_{11} + 2B_{12})P_R^2 = 0,$$

which can only be satisfied if $B_{11}+2B_{12}=0$ and $C_{11}+2C_{12}=0$. We therefore write $B_{11}=-2B_{12}=B$ and write out the equations for the strain in the other phases and obtain

$$\alpha(T_0-T)(c_{11}+2c_{12})+(c_{11}+C_{11}P_T^2)_T x + 2c_{12}y + BP_T^2 = 0,$$

$$\alpha(T_0-T)(c_{11}+2c_{12})+c_{12}x + (c_{11}+c_{12}+C_{12}P_T^2)_T y - \frac{BP_T^2}{2} = 0$$

for the tetragonal phase with x the polar axis and P_T the spontaneous polarization, while for the orthorhombic phase taking $P_x=P_y=P_0$ and $P_z=0$ with ${}_0x{}_x={}_0y{}_y\neq{}_0z{}_z$ we have

$$\alpha(T_0-T)(c_{11}+2c_{12})+[c_{11}+c_{12}+(C_{11}+C_{12})P_0^2]{}_0x{}_x + c_{12}{}_0z{}_z + \frac{BP_0^2}{2} = 0,$$

$$\alpha(T_0-T)(c_{11}+2c_{12})+2c_{12}{}_0x{}_x + (c_{11}+2C_{12}P_0^2){}_0z{}_z - BP_0^2 = 0,$$

$$c_{44}{}_0x{}_y + DP_0^2 = 0.$$

Solving these for the strains,

$${}_Tx{}_z = {}_Ty{}_y = \alpha(T-T_0) + (2c_{12}+c_{11}+C_{11}P_T^2) \frac{P_T^2}{2R},$$

$${}_Tx{}_x = \alpha(T-T_0) - (2c_{12}+c_{11}+C_{12}P_T^2) \frac{P_T^2}{R},$$

$${}_Tx{}_y = {}_Ty{}_z = {}_Tz{}_x = 0.$$

$${}_0z{}_z = \alpha(T-T_0) + [c_{11}+2C_{12}+(C_{11}+C_{12})P_0^2] \frac{P_0^2}{R},$$

$${}_0x{}_x = {}_0y{}_y = \alpha(T-T_0) - (c_{11}+2c_{12}+2C_{12}P_0^2) \frac{P_0^2}{2R},$$

$${}_0x{}_y = -\frac{DP_0^2}{c_{44}}; \quad {}_0y{}_z = {}_0x{}_z = 0,$$

where

$$R = (c_{11}+c_{12}+C_{12}P_T^2)(c_{11}+C_{11}P_T^2) - 2c_{12}^2,$$

$$\div (c_{11}-c_{12})(c_{11}+2c_{12}),$$

while in the rhombohedral phase we have simply

$${}_Rx{}_x = {}_Ry{}_y = {}_Rz{}_z = \alpha(T-T_0); \quad {}_Rx{}_y = {}_Ry{}_z = {}_Rz{}_x = -\frac{DP_R^2}{c_{44}}.$$

The volume above -90° C. is given to a second approximation by

$$V = 3\alpha(T-T_0) + \frac{(C_{11}-C_{12})BP^4}{c_{11}-c_{12}} + V_{T_0},$$

while below -90° C.

$$V = 3\alpha(T-T_0) + V_{T_0}.$$

To the same order above 90° C.

$$x_x - y_y = \frac{2BP^2}{c_{11}-c_{12}}.$$

Hence

$$V = 3\alpha(T - T_0) - \frac{1}{4}B^2(c_{11} - c_{12})(C_{11} - C_{12})(x_x - y_y)^2 + V_{T_0}^{\frac{1}{2}}$$

and

$$V/3 = \alpha(T - T_0) - \frac{1}{12}B^2(c_{11} - c_{12})(C_{11} - C_{12})(x_x - y_y)^2 + V_{T_0}.$$

The curve given by this equation is shown by the crosses on fig. 3 (a) and it is seen that it follows the experimental curve very closely and predicts a volume change at 120° C. of 0.02 Å³, the coefficient (C₁₁ - C₁₂) being chosen so as to obtain a fit at 0° C.

Substituting the values of x_x back into the expression for A the strain only contributes to the terms in P⁴ and P⁶ and thus the temperature dependent term in P² is unaffected. We obtain for the three phases

$$\begin{aligned} {}_T A &= \chi P_T^2 + P_T^4 \left[\frac{b}{4} - \frac{3B^2}{4(c_{11} - c_{12})} - \alpha B(C_{11} - C_{12}) \frac{c_{11} + 2c_{12}}{c_{11} - c_{12}} \right] + \dots P^6, \\ {}_O A &= 2\chi P_O^2 + P_O^4 \left[\frac{b}{2} - \frac{3B^2}{4(c_{11} - c_{12})} - \frac{D^2}{2c_{44}} - \alpha B(C_{11} - C_{12}) \frac{c_{11} + 2c_{12}}{c_{11} - c_{12}} \right] + \dots P^6, \\ {}_R A &= 3\chi P_R^2 + P_R^4 \left(\frac{3b}{4} - \frac{3D^2}{2c_{44}} \right) + \dots P^6. \end{aligned}$$

The result of the strain is to make the transition of the first kind if the total coefficient of P⁴ is negative (Devonshire 1949) and raises the transition temperature T. Experimentally T' - T₀ = 10° C. at 120° C. (Rushman 1946) and the strain energy is therefore quite small compared with kT and the transition temperature for the completely clamped crystal should be 110° C. and the reduction of the Curie temperatures to -5° C. and -90° C. largely due to the reduction of polarization perpendicular to the polar axes caused by the ion shifts.

The experimental result that the polarization is independent of the volume is unexpected for it implies that the shape of the potential wells of the ions are unaffected by a uniform expansion. Expanding the polarization in terms of the strain we have found

$$P_x^2 = a_1 x + a_2 x^2 - \frac{a_1}{2}(y + z) - \frac{a_2}{2}(y^2 + z^2),$$

a_1 and a_2 being proportional to B and C₁₁ - C₁₂. Now we may take x and y as proportional to the changes in the Ti-O and Ba-O distance d_{TiO} , d_{BaO} respectively. Thus

$$P^2 \propto \left(\frac{d_{TiO}}{d_{BaO}} - \frac{1}{\sqrt{(2)}} \right)$$

expresses our results. This result can be accounted for if the strain accompanying the polarisation alters the potential well of the oxygen ion only, for if the Ti ion vibration was affected we should expect P² to be proportional to δd_{TiO} , and hence a volume increase at -90° C. in contradiction to experiment. This is in qualitative agreement with our previous ideas for the large degree of freedom of the oxygen ion depends jointly on the small Ba-O distance and large TiO distance, and a uniform

expansion will not affect this, for thereby the ratio $d_{\text{TiO}}/d_{\text{BaO}}$ is unchanged. The conclusion that $B_{11}+2B_{12}=0$ is unlikely to be exactly obeyed, but more accurate determinations of the expansion coefficient in the cubic phase are desirable to investigate this point in detail, and also to measure the predicted volume change at 120°C. of 0.02 \AA^3 .

On fig. 3(c) is plotted the spontaneous polarization as a function of temperature calculated from the strain, an average value of 14×10^{-6} coulombs/cm.² being assumed for its value at 18°C. , this being the mean of the results of Hulm (1947) and Matthias and Von Hippel (1948). The change of P at -90°C. is rather approximate and is calculated from the associated change in the angle of shear ($\delta-\gamma$), the total resultant polarizations in the $[110]$ and $[111]$ directions also being shown. We see that the temperature saturation value along the cube axes is nearly reached at 0°C. in the tetragonal phase and the saturation value in the rhombohedral phase is somewhat less, in agreement with our earlier discussion.

* * * * *

CONCLUDING REMARKS.

The preceding discussion is intended to show that, granted the assumptions underlying the Lorentz equation, there is a short range dipole interaction which is of probably greater magnitude than the Lorentz term itself, and also that this interaction accounts for the three observed transitions in a simple manner. Some difficulties are still left however which will be briefly discussed.

The short range effect increases the local field so as to give an effective Lorentz factor (β) of $3 \times 4\pi/3$. On any purely dipole system it appears that in the Curie Weiss law $\epsilon=K/T-T_0$ with the usual nomenclature gives $\beta=GT_0/K$ where G is a numerical constant not very different from unity, depending on the dipole system considered. Now $T_0/K \sim 10^{-2}$ making $B=10^{-2}G$ and thus abnormally low. We are left with two alternatives, either to accept a low value of B or to examine again the validity of the treatment by permanent dipole methods. The first alternative was adopted by Mason and Matthias and satisfactory results obtained assuming $B=0.134$, G being 4π with their six Ti-position model. No definite explanation of the low value is given though it is suggested as being due to the displacement of the Ti ion towards an oxygen ion making the Lorentz assumptions of cubic symmetry invalid. This is equivalent to supposing a short range force acting in opposition to the Lorentz field whose origin is not known. A low value of β therefore implies that these two opposing fields are almost equal, for β is reduced from 4.2 to 0.134. This is most unlikely for this value of β is fundamental to this theory of ferroelectrics as it determines the cooperative field, and the implicit postulation of the opposition field makes the value of the net field dependent on the difference of two almost equal quantities.

A possible solution of this difficulty is to postulate that the dipole moment is of the form $\mu_0+\mu_1T$, physically meaning that the potential well of the ions are of the form $\alpha x^2+\beta x^4$, α and β being related to μ_0 and

μ_1 . This has been developed by Devonshire and again leads to the Curie point being determined by the difference of two almost equal quantities, a result which still seems rather arbitrary.

It is possible that quantitative statistical treatment of the dipole array suggested may solve the difficulty, for the inclusion of a short range force would be likely to increase the slope of the reciprocal susceptibility curve in agreement with experiment. A treatment analogous to Bethe's (1935) method for order-disorder in alloys may be applicable, short range interaction giving rise to long range order at the Curie point, but this brings us to a second serious difficulty, the propagation of short range order perpendicular to the polarization.

In ferromagnetism the exchange interaction always makes the parallel alignment of dipoles the lowest energy state independent of their mutual orientation. In electrical dipole interaction, however, this is not true for if $\theta > \tan^{-1}\sqrt{2}$ the antiparallel orientation has lowest energy. Considering only short range interaction the antiparallel array of adjacent chains of dipoles has lowest energy for most simple types of arrangement (Sauer 1940) including the one discussed for BaTiO_3 . This however cannot occur in practice for there is no resultant polarization with this system. The treatment of cooperative effects in one dimension only does not give rise to long range order (Bethe 1935) so that for a transition to occur there must be an aligning interaction perpendicular to the dipole axis, which is not apparent in the idealized simple array for BaTiO_3 considered here. The inclusion of the long range Lorentz term which results from surface charges collected to neutralize the polarization does not help, for all arrangements have the same energy as far as this interaction is concerned.

The recent developments of the theory of the polarizability of ionic crystals (Szigeti 1949) may help in this respect, for throughout this discussion electronic polarizability has been treated similarly to that of ionic shifts and this polarization is of considerable importance in BaTiO_3 (cf. Hulm 1947, Fröhlich 1949).

ACKNOWLEDGMENTS.

We are indebted to our colleagues Dr. A. F. Devonshire and Mr. H. J. Wellard for very valuable suggestions, to the latter for the production of suitable crystals, and also to the Electrical Research Association for financial assistance in the purchase of equipment. One of us (P.V.) also wishes to acknowledge financial assistance from the Department of Scientific and Industrial Research.

REFERENCES.

- MEGAW, H. D., 1947, *Proc. Roy. Soc. A*, **189**, 261.
HARWOOD, M. G., 1949, *Journ. Sci. Instr.*, **26**, 137.
KAY, H. F., 1948, *Acta Cryst.*, **1**, 229.
KAY, H. F., WELLARD, H. J., and VOUSDEN, P., 1949, *Nature*, **163**, 636.
ROBERTS, S., 1947, *Phys. Rev.*, **71**, 890.
ROBERTS, S., 1949, *Phys. Rev.*, **75**, 989.

- MERZ, M. J., 1948, *Phys. Rev.*, **75**, 687.
 RUSHMAN, D. F., and STRIVENS, M. A., 1946, *Trans. Far. Soc.*, **42**, 231.
 UBBELOHDE, A. R., and WOODWARD, H., 1947, *Proc. Roy. Soc. A*, **188**, 358.
 MASON, W. P., and MATTHIAS, B. T., 1948, *Phys. Rev.*, **74**, 1622.
 HULM, F., 1947, *Nature*, **160**, 126.
 DEVONSHIRE, A. F. In this issue.
 SAUER, J. A., 1940, *Phys. Rev.*, **57**, 142.
 BETHE, H. A., 1935, *Proc. Roy. Soc. A*, **150**, 552.
 SZIGETI, B., 1949, *Trans. Far. Soc.*, **45**, 155.
 DANIELSON, G. C., MATTHIAS, B. T., and RICHARDSON, J. M., 1948, **74**, 986.
 MOTT, N. F., and GURNEY, R. W., 1940, *Electronic Processes in Ionic Crystals*,
 (Oxford : University Press).
 FRÖHLICH, H., 1949, *Theory of Dielectrics*. (Oxford : University Press).

XCVI. *Theory of Barium Titanate.*—Part I.

By A. F. DEVONSHIRE,
 H. H. Wills Physical Laboratory, University of Bristol*.

[Received July 26, 1949.]

SUMMARY.

The theory of the dielectric and crystallographic properties of barium titanate is considered. By expanding the free energy as a function of polarization and strain and making reasonable assumptions about the coefficients, it is found possible to account for the various crystal transitions. Calculations are made of the dielectric constants, crystal strains, internal energy, and self polarization as functions of temperature. Finally relations are obtained between the coefficients in the free energy and the ionic force constants. These are used to estimate some of the coefficients which are not completely determined by experimental data.

§1. INTRODUCTION.

In the last few years the properties of a number of substances known as ferroelectrics or seignette-electrics have been much studied. At present three groups of these substances are known, typical members of the three groups being Rochelle salt, potassium dihydrogen phosphate, and barium titanate. All these substances have certain properties in common. At sufficiently high temperatures their properties are normal, though the dielectric constants are usually rather high. As the temperature falls the dielectric constant increases and reaches a peak at a transition temperature. At this temperature there is a change of crystal form to one of lower symmetry. Below this temperature each crystal breaks up into domains and there is clear evidence that these domains are polarized. The substance shows the properties of hysteresis and saturation that one would expect from such a structure. There may be lower

* Communicated by the Author.

transition temperatures. Rochelle salt becomes normal again at still lower temperatures, and barium titanate has two lower transition temperatures at which there are further changes of crystal form. The crystal changes are always small, the shears involved being usually less than a degree. There are also small specific heat changes at the transition temperature.

In this paper we shall consider only the most recently discovered group of ferroelectrics, the third, and in particular barium titanate. This is the only known pure substance in the group, though solid solutions of barium titanate with lead or strontium titanate show similar properties.

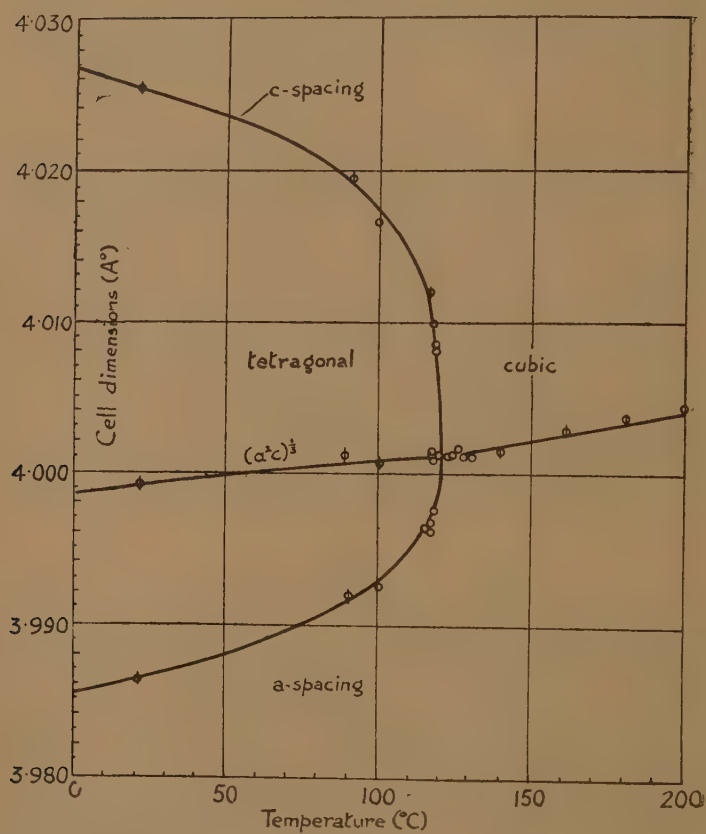
We shall first review the experimental evidence on the crystal structure, specific heat, saturation polarization and dielectric constant for small fields. We shall then discuss the theory of the substance, first on a phenomenological basis and then in terms of a molecular model. This will include a discussion of theories already put forward. We shall not consider the dielectric constant at large fields, nor any of the time-dependent phenomena, such as hysteresis or the dependence of dielectric constant on frequency of field. We hope to deal with some of these in a later paper.

§2. CRYSTAL STRUCTURE.

Above the transition temperature barium titanate BaTiO_3 has a cubic structure. The barium ions lie at the corners of a cubic lattice, the titanium ions at the body centres, and the oxygen ions at the face centres. Below 120°C . it was shown by Megaw (1946) that the substance becomes tetragonal; one of the axes (usually taken to be the c -axis) becomes lengthened, and the other two shortened. The axial lengths as a function of temperature are shown in fig. 1. It will be seen that the change appears to set in rather abruptly. This was verified by Harwood, Popper and Rushman (1947), who showed that in a given crystallite c/a changed discontinuously from 1 to 1.005. There is a range of a few degrees, however, in which the substance is a mixture of cubic and tetragonal forms. Optical studies by Kay (1948) Matthias and Von Hippel (1948) and Blättner, Kanzig and Merz (1949) show that each crystal has broken up into a number of domains. In the simplest type the domains are arranged in the way shown in fig. 2. The domain extends from one face to a parallel one and the directions of the tetragonal axes are as shown.

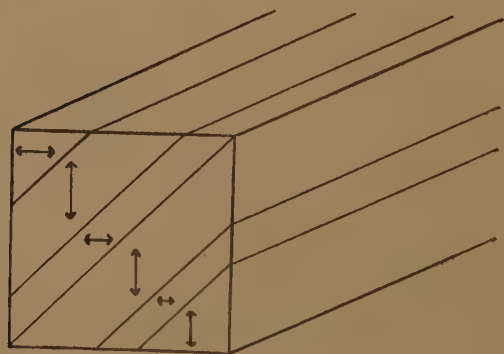
More recent optical and X-ray studies by Kay, Vousden and Wellard (1949) show that there is a further transition at about -10°C . Below this temperature the crystal is orthorhombic. At the transition the c -parameter shortens slightly and the a -parameter increases so that the two become equal and there is a shear of about $14'$ in the ca plane. The polar axis, which was formerly in the c -direction is now along a diagonal in the ca plane. The above authors also report that there is a further transition at about -70°C . Below this temperature the crystal is probably rhombohedral with the polar axis along the $[111]$ direction.

Fig. 1.



Lattice spacing of BaTiO₃ as a function of temperature (Megaw 1946).

Fig. 2.



§3. SPECIFIC HEAT AND SPONTANEOUS POLARIZATION.

It has been found by Wul (1946) and also by Harwood, Popper and Rushman (1947) that there is a hump in the specific heat curve in the neighbourhood of 120° C. Blattner, Kanzig and Merz (1949) found a hump in the neighbourhood both of 120° C. and of 0° C. The additional specific heat is only a small fraction of the normal specific heat so it is difficult to separate the two, but Blattner, etc. found that the total additional heat was about 47 cal./mole at the higher transition, and about 16 cal./mole at the lower transition. The other authors' results are not stated in their papers, but estimating very roughly from their curves the total additional heat appears to be about 20 or 30 cal./mole.

There is only a limited amount of evidence on the spontaneous polarization. A field which is strong enough to orientate all the domains in the same direction will produce considerable induced polarization since the domains are themselves highly polarizable. It is difficult, therefore, to separate the spontaneous from the induced polarization. For sufficiently strong fields, however, the polarization becomes a linear function of field strength, and by extrapolating back to zero field Hulm (1947) estimated that the spontaneous polarization at room temperature was about 16 microcoulombs per cm.². Matthias and Von Hippel (1948) estimated it to be about 12 microcoulombs/cm.². Unfortunately we do not know the spontaneous polarization as a function of temperature, but Hulm (1947) has given the total polarization for large fields as a function of temperature. As the temperature falls it rises rather rapidly in the neighbourhood of 120° C. and then increases slowly. The measurements do not go below 0° C.

§4. DIELECTRIC CONSTANT FOR SMALL FIELDS.

Numerous measurements of the dielectric constant for small fields have been made, usually with alternating current. Papers by various Russian authors have been summarized by Wul (1946). Jackson and Reddish (1945) and Rushman and Strivens (1946) have studied the effect of varying the composition. Von Hippel, Brockenridge, Chesley and Tisza (1946) have published many measurements for various compositions, field strengths, and frequencies. All the authors agree that above the highest transition temperature the dielectric constant obeys a Curie-Weiss law; that is

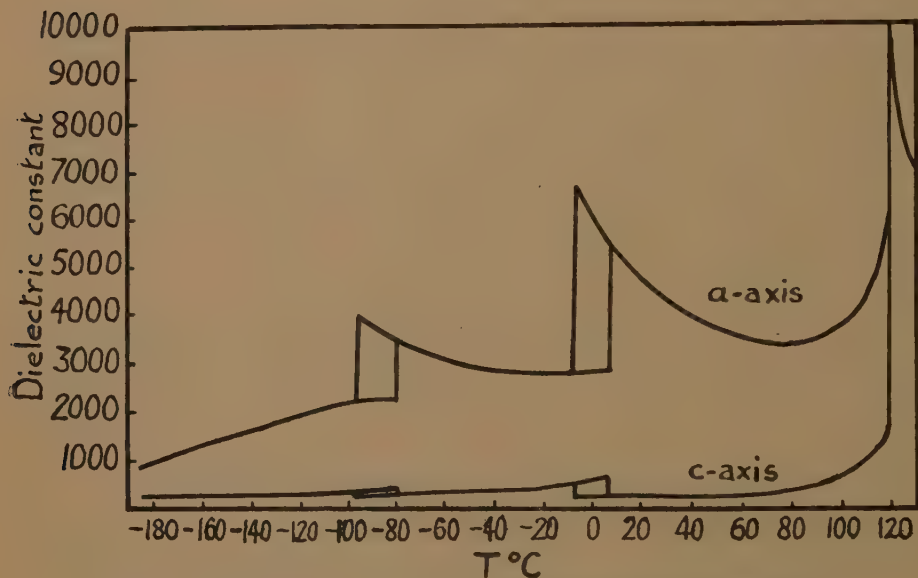
$$\epsilon = \frac{A}{T - T_a} + \epsilon_0,$$

where ϵ_0 may or may not be zero and A is very large, about 10^5 degrees. Below this temperature the dielectric constant drops to about 1000–2000, and as the temperature decreases remains constant or slowly decreases. There are small peaks at lower temperatures. These almost certainly correspond to the lower crystal transitions. All the above measurements were made with powders.

Now, above the highest transition temperature the substance has complete cubic symmetry and therefore a single dielectric constant;

but below this temperature the dielectric constants along the different axes will be different and experiments on powders will only give a mean value. Matthias and Von Hippel (1948) have carried out experiments on single crystals; but even a single crystal may contain several differently orientated domains. However, they made measurements on a crystal in which all the domains had been made parallel by applying a strong electric field, and they found a dielectric constant of 500 along the c -axis and 1700 along the a -axis. Mason and Matthias made measurements on single domain crystals and found dielectric constants along the a -axis

Fig. 3.

Dielectric constant of BaTiO₃ single domain crystals (Merz 1949).

of the order of 10^5 . Merz, however, found more moderate values. His results are shown in fig. 3. The discontinuities at the lower transition temperatures are very well marked.

§5. THEORIES.

Suggestions put forward to account for the behaviour of barium titanate are usually based on the assumption that the titanium ion can move rather easily. Miss Megaw (1946) has pointed out that if the ions are all assumed to have the Goldschmidt radii they cannot be fitted together exactly to make a cubic structure but that they have a certain amount of free space. Rushman and Strivens (1946) have developed this idea by suggesting that the titanium ions are slightly displaced from the symmetrical position and therefore form dipoles which can rotate. It is known that such a set of dipoles have a Curie temperature above which they rotate freely, but below which they are parallel to one another. This latter state, of course, corresponds to the ferroelectric region.

Mason and Matthias have developed in considerable detail a rather different model. They assume that the titanium ion has six equilibrium positions slightly displaced in the axial directions from the symmetrical one. In the cubic state the ions occupy the positions at random. In the tetragonal state they occupy mainly positions along one axis. The transition from cubic to tetragonal symmetry therefore corresponds to a transition from a disordered to an ordered state. By assuming suitable values for the ionic shift, the polarizability of the rest of the material, and the Lorentz factors, they have accounted for the dielectric constant in the cubic region, and the two dielectric constants in the tetragonal region. They have also considered hysteresis and frequency effects.

The greater part of this paper considers the theory of barium titanate in a phenomenological way. We expand the free energy in terms of the strains and polarization of the crystal, use certain properties of the crystal to determine the coefficients, and then predict other properties. Results obtained in this way are of course, independent of any atomic model. This method has been applied to Rochelle salt with considerable success by Mueller (1940), but owing to the very different symmetry of barium titanate we cannot make any direct use of his results. Mason (1948) has applied this method to the electrostrictive effect in barium titanate; Ginsburg (1946) has applied it to the highest transition, but only a limited amount of experimental evidence was available when his paper was published. We have been able to show that by assuming reasonable values for the coefficients we can explain the successive transitions through the cubic, tetragonal, orthorhombic and rhombohedral forms. In determining the coefficients we use the observed transition temperatures, the value of the dielectric constant in the cubic region, and the observed strain and saturation polarization at a single temperature in the tetragonal region. We are then able to predict the various dielectric constants in the tetragonal, orthorhombic and rhombohedral regions. We also predict values for the strain and saturation polarization in these regions. It is not possible to verify all these predictions as many of these quantities have not yet been observed.

Finally we consider an atomic model for barium titanate. Following the method used by Born for ionic crystals we treat the ions mainly as point centres of force, though also taking into account their polarizability. After checking the force constants by calculating the interatomic distance we make an estimate of the elastic constants. In conjunction with the results already obtained this enables us to calculate the electrostrictive constants. We then attempt to calculate directly the field in which each ion moves, the other ions being in the symmetrical positions. These calculations clearly indicate that the stable position of the titanium ion is the symmetrical one. It seems just possible that the oxygen ions might have two unsymmetrical equilibrium positions displaced towards the nearest Ti ions, but the calculations are not accurate enough to make a definite prediction. We have, however, assumed that an ion moving individually has only one position of equilibrium, and that the spontaneous polarization

is caused by the Lorentz field, in other words, it is a cooperative effect, since this field only exists when ions of one sign move together.

The Lorentz field requires some consideration since it plays an important part in the theory of ferroelectrics. When a body becomes polarized an ion or dipole in the body will experience a force due to the polarization of the rest of the body. We may assume this force to be proportional to the polarization and put it equal to βP where β is the Lorentz factor. Owing to the slow fall off with distance of the dipole force it is a "long range" one, that is dipoles in distant parts of the body have an appreciable effect. Hence for an insulated body β is dependent on external shape and reaches its maximum value for a needle or plate polarized parallel to a long axis. But for an uninsulated body β always has this value whatever the external shape. A maximum value of β corresponds to a minimum polarization energy since this is $-\frac{1}{2}\beta P^2$ per unit volume, and hence the uninsulated body will always collect surface charges in such a way as to make β a maximum, say β_m . For point dipoles in a cubic or random array β_m can be shown to be $4\pi/3$ (Fowler 1936). If the array is regular but not cubic β_m will vary somewhat with direction though its mean value will still be $4\pi/3$. If the dipoles are not point dipoles β_m will have a different value.

Mason and Matthias have found it necessary to assume that β_m is slightly different for the tetragonal and the other directions and also varies slowly with the temperature. These assumptions are reasonable, but the values they have found it necessary to assume for β_m are near 0.10. This differs considerably from the theoretical value of 4.19 even taking into account the fact that the dipoles are far from being point dipoles. Physically it means that the cooperative effect between the dipoles is rather small.

We have assumed that β_m has the value $4\pi/3$, and that the Lorentz field approximately balances the short range restoring force. By taking into account thermal vibrations, including anharmonic terms, we are able to show that the restoring force increases with temperature, and hence explain the existence of a transition temperature.

For the properties we are dealing with in this paper there is no important difference between the predictions of the two models. The Mason Matthias model would need a little modification to account for the lower transitions, but this could certainly be done. There are bigger differences in other phenomena, but these will be dealt with in a later paper.

§6. PHENOMENOLOGICAL THEORY: INTRODUCTION.

We shall consider the substance as a strained cubic crystal. All the changes from cubic symmetry are small, so this is quite legitimate. We shall use the notation given by Cady in his "Piezoelectricity", as far as possible. For barium titanate, however, it is necessary to consider higher order terms in the free energy than any used by Cady, so we shall have to introduce some new notation.

Now the free energy of a crystal can be expressed in several different forms. We can take as our independent variables polarization and stress,

polarization and strain, field and stress or field and strain. We shall start by expressing the free energy as a function of polarization and stress with the stresses equated to zero. We then have

$$A = \frac{1}{2}\chi' \{P_x^2 + P_y^2 + P_z^2\} + \frac{1}{4}\xi'_{11} \{P_x^4 + P_y^4 + P_z^4\} + \frac{1}{2}\xi'_{12} \{P_y^2 P_z^2 + P_z^2 P_x^2 + P_x^2 P_y^2\} \\ + \frac{1}{6}\zeta' \{P_x^6 + P_y^6 + P_z^6\}. \quad (6.1)$$

The zero of free energy is taken to be that of the unpolarized, unstressed crystal. All the terms of the second and fourth orders are included and some of the terms of the sixth order. We shall find it necessary to take into account all these terms to account for the behaviour of the crystal. The derivatives of A with respect to P_x give the field-components for the free, that is unstressed, crystal. Hence we have

$$E_x = \chi' P_x + \xi'_{11} P_x^3 + \xi'_{12} P_x (P_y^2 + P_z^2) + \zeta' P_x^5. \quad (6.2)$$

In the absence of a field the right-hand side of (6.2) must be zero and for stability A must be a minimum.

We shall find that we can account satisfactorily for the observed facts if we assume that ζ' and ξ'_{12} are positive, ξ'_{11} is negative, and χ' is a decreasing function of temperature which passes through a zero value in the neighbourhood of the upper transition temperature. Since χ' is the reciprocal susceptibility for zero polarization we can obtain its value from experiment directly above the upper transition temperature where there is zero polarization for zero field. We find, in fact, that it is a linear decreasing function of temperature which, if extrapolated, passed through zero a little below the transition temperature. With these assumptions A will clearly be always positive for sufficiently large χ' , that is, at sufficiently high temperatures. The minimum value of A will then correspond to zero polarization. For χ' small or negative, however, the minimum value of A will correspond to a finite polarization. The second order term is independent of direction. But for a given resultant polarization the fourth order term has its minima along the axes and the sixth order term has its minima along the diagonal directions. Hence as the temperature falls and the magnitude of the polarization increases the direction will be likely to change from an axial to a diagonal one.

For zero field, equation (6.2) and the similar equations become

$$\left. \begin{aligned} P_x = 0, \text{ or } \zeta' P_x^4 + \xi'_{11} P_x^2 + \xi'_{12} (P_y^2 + P_z^2) + \chi' &= 0, \\ P_y = 0, \text{ or } \zeta' P_y^4 + \xi'_{11} P_y^2 + \xi'_{12} (P_z^2 + P_x^2) + \chi' &= 0, \\ P_z = 0, \text{ or } \zeta' P_z^4 + \xi'_{11} P_z^2 + \xi'_{12} (P_x^2 + P_y^2) + \chi' &= 0. \end{aligned} \right\} \quad (6.3)$$

There are four sets of solutions of these equations which may correspond to minima of A , namely

$$\left. \begin{aligned} P_x = P_y = P_z &= 0, & (a) \\ P_x = P_y = 0, \zeta' P_z^4 + \xi'_{11} P_z^2 + \chi' &= 0, & (b) \\ P_x = 0, P_y = P_z, \zeta' P_z^4 + (\xi'_{11} + \xi'_{12}) P_z^2 + \chi' &= 0, & (c) \\ P_x = P_y = P_z, \zeta' P_z^4 + (\xi'_{11} + 2\xi'_{12}) P_z^2 + \chi' &= 0, & (d) \end{aligned} \right\} \quad (6.4)$$

with, of course, the corresponding solutions obtained by interchanging P_x , P_y , and P_z . The corresponding values of A are given by

$$\left. \begin{aligned} A &= 0, & (a) \\ A &= \frac{1}{6}\zeta'P_z^6 + \frac{1}{4}\xi'_{11}P_z^4 + \frac{1}{2}\zeta'P_z^2, & (b) \\ A &= \frac{1}{3}\zeta'P_z^6 + \frac{1}{2}(\xi'_{11} + \xi'_{12})P_z^4 + \chi'P_z^2, & (c) \\ A &= \frac{1}{2}\zeta'P_z^6 + \frac{3}{4}(\xi'_{11} + 2\xi'_{12})P_z^4 + \frac{3}{2}\chi'P_z^2. & (d) \end{aligned} \right\} \quad . \quad . \quad (6.5)$$

The necessary conditions for A to be a minimum are

$$\left. \begin{aligned} \chi' &> 0, & (a) \\ \chi' + \xi'_{12}P_z^2 &> 0, \quad \frac{1}{2}\xi'_{11} + \zeta'P_z^2 &> 0, & (b) \\ \chi' + 2\xi'_{12}P_z^2 &> 0, \quad \frac{1}{2}\xi'_{11} + \zeta'P_z^2 &> 0, & (c) \\ & \frac{1}{2}\xi'_{11} + \zeta'P_z^2 &> 0. & (d) \end{aligned} \right\} \quad . \quad . \quad . \quad (6.6)$$

It is possible for A to have more than one minimum, and we then have to determine which is the least. If we plot the values of A given by (6.4) and (6.5) as a function of χ' , taking ζ' and ξ' to be constant then we find that for χ' positive and large enough the least minimum of A is zero, but as χ' decreases the least minimum is successively given by equations (b), (c) and (d). When the minimum given by (b) is equal to that given by (a) we have

$$\frac{1}{6}\zeta'P_z^4 + \frac{1}{4}\xi'_{11}P_z^2 + \frac{1}{2}\chi' = 0, \quad . \quad . \quad . \quad . \quad . \quad (6.7)$$

and from (6.4 b) we have

$$\zeta'P_z^4 + \xi'_{11}P_z^2 + \chi' = 0. \quad . \quad . \quad . \quad . \quad . \quad (6.8)$$

Let us denote by χ'_0 the value of χ' which satisfies these equations and by P_1 the corresponding value of P_z . Then we have

$$\zeta' = 3\chi'_0/P_1^4, \quad . \quad . \quad . \quad . \quad . \quad (6.9)$$

$$\xi'_{11} = -4\chi'_0/P_1^2. \quad . \quad . \quad . \quad . \quad . \quad (6.10)$$

If we assume that ζ' and ξ'_{11} are constants, independent of temperature, and therefore always given by (6.9) and (6.10), and if we also put

$$P_z^2 = zP_1^2, \quad P_x^2 = xP_1^2, \quad P_y^2 = yP_1^2, \quad . \quad . \quad . \quad . \quad . \quad (6.11)$$

$$\chi' = t\chi'_0, \quad . \quad . \quad . \quad . \quad . \quad (6.12)$$

$$\xi'_{12} = -\alpha\xi'_{11}, \quad . \quad . \quad . \quad . \quad . \quad (6.13)$$

then equations (6.4) and (6.5) become

$$\left. \begin{aligned} x &= y = z = 0, & (a) \\ x &= y = 0, \quad 3z^2 - 4z + t = 0, & (b) \\ x &= 0, \quad y = z, \quad 3z^2 + 4(-1 + \alpha)z + t = 0, & (c) \\ x &= y = z, \quad 3z^2 + 4(-1 + 2\alpha)z + t = 0, & (d) \end{aligned} \right\} \quad . \quad . \quad (6.14)$$

and

$$\left. \begin{aligned} A &= 0, \\ A &= \chi'_0 P_1^2 \left\{ \frac{1}{2} z^3 - z^2 + \frac{1}{2} z t \right\}, \\ A &= \chi'_0 P_1^2 \{ z^3 + 2(-1 + \alpha) z^2 + z t \}, \\ A &= \chi'_0 P_1^2 \left\{ \frac{3}{2} z^3 + 3(-1 + 2\alpha) z^2 + \frac{3}{2} z t \right\}. \end{aligned} \right\} \dots \dots (6.15)$$

The equations are now in a convenient form for calculation, since they are now in the form of a relation between two variables z and t with a single parameter α . The other constants of the equation enter only as scale factors. The variable t is related to the temperature, and the relation is probably approximately linear. We have seen this is the case experimentally for t positive, since χ' is proportional to t . In fig. 4 we plot z and $A/\chi'_0 P_1^2$ as functions of t for α equal to 1.2. As already stated the minimum value of A is given in turn by equations (a), (b), (c) and (d). The value of z will therefore change discontinuously at three transition temperatures as z is given in turn by equations (a), (b), (c) and (d). The polarization, which was originally zero, will in turn point along a cube edge, a face diagonal, and a body diagonal. The corresponding effect on the crystal symmetry will be to change it from the original cubic successively to tetragonal, orthorhombic and rhombohedral. In the Schönflies notation the symmetry of the crystal will be in turn O_h , C_{4v} , C_{2v} and C_{3v} .

§7. PHENOMENOLOGICAL THEORY: DIELECTRIC CONSTANT.

In the high temperature region where the crystal has cubic symmetry the dielectric constant is independent of direction and is given by

$$\epsilon = 1 + 4\pi/\chi', \dots \dots \dots (7.1)$$

but in the temperature regions where the crystal has less symmetry the dielectric constant is no longer independent of direction and has to be described by a tensor. It is, in practice, more convenient to work in terms of the susceptibility η_{rs} and the reciprocal susceptibility χ_{rs} . For small fields and polarization they are defined by the relations

$$\begin{aligned} E_x &= \chi_{11} P_x + \chi_{12} P_y + \chi_{13} P_z, \\ &\dots \dots \dots \\ P_x &= \eta_{11} E_x + \eta_{12} E_y + \eta_{13} E_z, \quad \dots \dots \dots (7.2) \\ &\dots \dots \dots \end{aligned}$$

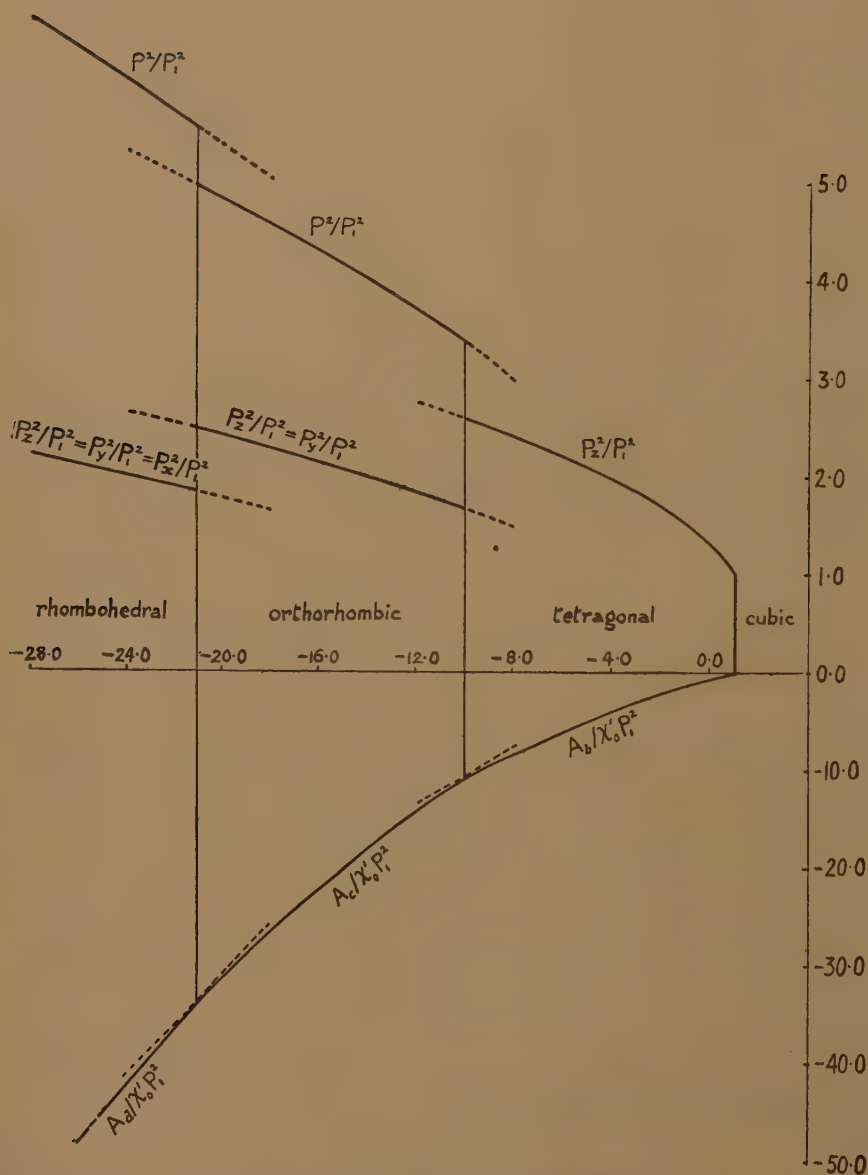
and are connected by the equation

$$\eta_{rs} = \chi_{rs} / \Delta, \quad \dots \dots \dots (7.3)$$

where Δ is the determinant of the χ 's and X_{rs} is the minor of χ_{rs} in that determinant. When the relation between field and polarization is no longer linear we define χ by the equation

$$\chi_{sr} = \chi_{rs} = \frac{\partial E_r}{\partial P_s} = \frac{\partial E_s}{\partial P_r} = \frac{\partial^2 A}{\partial P_r \partial P_s} \dots \dots \dots (7.4)$$

Fig. 4.



Free energy and polarization of BaTiO_3 as a function of temperature (theoretical).

Hence from (6.1) we have

$$\left. \begin{aligned} \chi_{zz} &= \chi' + 3\xi'_{11}P_z^2 + \xi'_{12}(P_x^2 + P_y^2) + 5\xi'_{22}P_z^4 \\ \chi_{xy} &= \xi'_{12}P_xP_y, \end{aligned} \right\} \dots \dots (7.5)$$

with similar expressions for χ_{xx} etc. Using the equations (6.9) to (6.13) we have

$$\left. \begin{aligned} \chi_{zz} &= \chi'_0 \{t - 12z + 4\alpha(x+y) + 15z^2\}, \\ \chi_{xy} &= 4\alpha\chi'_0 x^{\frac{1}{2}}y^{\frac{1}{2}}. \end{aligned} \right\} \dots \dots (7.6)$$

If we now consider our four cases in turn we find that

$$\left. \begin{aligned} \chi_{xx} &= \chi_{yy} = \chi_{zz} = \chi' = r\chi'_0, \\ \chi_{yz} &= \chi_{zx} = \chi_{xy} = 0, \end{aligned} \right\} (a) \\ \left. \begin{aligned} \chi_{xx} &= \chi_{yy} = \chi'_0(t + 4\alpha z), \\ \chi_{zz} &= \chi'_0(t - 12z + 15z^2), \\ \chi_{yz} &= \chi_{zx} = \chi_{xy} = 0, \end{aligned} \right\} (b) \\ \left. \begin{aligned} \chi_{xx} &= \chi'_0(t + 8\alpha z), \\ \chi_{yy} &= \chi_{zz} = \chi'_0(t - 12z + 4\alpha z + 15z^2), \\ \chi_{xy} &= \chi_{zx} = 0, \quad \chi_{yz} = \chi'_0 4\alpha z, \end{aligned} \right\} (c) \\ \left. \begin{aligned} \chi_{xx} &= \chi_{yy} = \chi_{zz} = \chi'_0(t - 12z + 8\alpha z + 15z^2), \\ \chi_{yz} &= \chi_{zx} = \chi_{xy} = \chi'_0 4\alpha z. \end{aligned} \right\} (d) \end{aligned} \right\} \dots \dots (7.7)$$

For cases (a) and (b) χ_{xx} , χ_{yy} and χ_{zz} are the principal values of the reciprocal susceptibilities, but for (c) and (d) they are not. For case (c) the principal directions are the x -axis and the two face diagonals in the yz -plane. The reciprocal susceptibilities in these directions are

$$\left. \begin{aligned} \chi_{xx} &= \chi'_0(t + 8\alpha z), \\ \chi_{\beta\beta} &= \chi_{zz} + \chi_{yz} = \chi'_0(t - 12z + 8\alpha z + 15z^2), \\ \chi_{\gamma\gamma} &= \chi_{zz} - \chi_{yz} = \chi'_0(t - 12z + 15z^2). \end{aligned} \right\} \dots \dots (7.8)$$

$\chi_{\beta\beta}$ gives the reciprocal susceptibility in the direction of polarization. For case (d) the principal directions are a body diagonal and any direction perpendicular to it. The reciprocal susceptibilities are given by

$$\left. \begin{aligned} \chi_{\alpha\alpha} &= \chi_{zz} + 2\chi_{xy} = \chi'_0(t - 12z + 16\alpha z + 15z^2), \\ \chi_{\delta\delta} &= \chi_{zz} - \chi_{xy} = \chi'_0(t - 12z + 4\alpha z + 15z^2). \end{aligned} \right\} \dots \dots (7.9)$$

$\chi_{\alpha\alpha}$ gives the reciprocal susceptibility in the direction of polarization. The reciprocal susceptibility in all perpendicular directions is the same and is given by $\chi_{\delta\delta}$.

We must now compare our results with experiment and determine the parameters we have used. These are χ'_0 , $\beta_0 P_1$, and the relation between t and the temperature. Now in the non-ferroelectric region we have

$$\chi = t\chi'_0,$$

and experimentally χ is a linear function of temperature. We can therefore reasonably assume that t is a linear function of temperature throughout, and since by definition it is 1 at the upper transition temperature, which we shall call T_1 , we can conveniently write

$$t = (T - T_0)/(T_1 - T_0), \quad . \quad . \quad . \quad . \quad . \quad (7.10)$$

where T_0 is a parameter to be determined. Since T_0 is the temperature at which t and therefore χ vanishes we can determine it by extrapolating χ to zero. Unfortunately it is only slightly less than T_1 , and since the transition is spread over a few degrees this means that $T_1 - T_0$ is rather uncertain. From the measurements of Harwood, Popper and Rushman (1947) we estimate that T_0 is 118°C . and $T_1 - T_0$ about 10°C ., making T_0 128°C . Since

$$\chi = t\chi'_0 = \chi'_0(T - T_0)/(T_1 - T_0), \quad . \quad . \quad . \quad . \quad (7.11)$$

we can determine $\chi'_0/(T_1 - T_0)$ from the observed slope of χ plotted against temperature. The measurements of different observers agree fairly well and we find that

$$\chi'_0/(T_1 - T_0) = 1.0 \times 10^{-4} (\text{degrees})^{-1}, \quad . \quad . \quad . \quad . \quad (7.12)$$

and hence if we take $T_1 - T_0$ equal to 10° , then

$$\chi'_0 = 1.0 \times 10^{-3}. \quad . \quad . \quad . \quad . \quad . \quad (7.13)$$

P_1 can be determined by making the calculated saturation polarization agree with the observed value at a given temperature. If we take P_1 to be 10 microcoulombs/cm.², then from (6.11), (6.14) and (7.10) we can show that the saturation polarization is about 16 microcoulombs/cm.² at room temperature in agreement with Hulm's result (1947).

α is best determined by choosing it so that one of the lower transition temperatures agrees with the observed value. We find, in fact, that if we take α to be 1.2 both transition temperatures agree fairly well with the observed values.

If we substitute the above values in (6.9), (6.10) and (6.13) we find that

$$\xi'_{11} = -4.4 \times 10^{-12}, \quad . \quad . \quad . \quad . \quad . \quad (7.14)$$

$$\xi'_{12} = 5.3 \times 10^{-12}, \quad . \quad . \quad . \quad . \quad . \quad (7.15)$$

and

$$\zeta' = 3.7 \times 10^{-21}. \quad . \quad . \quad . \quad . \quad . \quad (7.16)$$

We can now plot the principal dielectric constants as a function of temperature. The dielectric constants in the principal directions are related to the reciprocal susceptibilities by the relation

$$\epsilon = 1 + 4\pi\chi^{-1}.$$

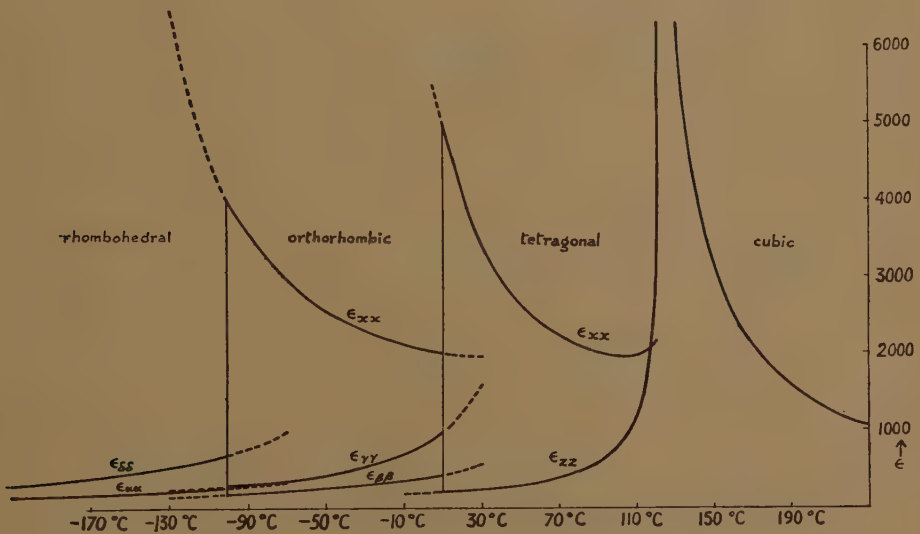
The results are shown in fig. 5. If we compare them with the results of Merz given in fig. 3 we see that the theoretical curves have the same form,

but there is no exact numerical agreement. In the tetragonal region our ϵ_{xx} and ϵ_{zz} correspond to Merz's ϵ_a and ϵ_c . In the orthorhombic region ϵ_a and ϵ_c are no longer principal dielectric constants; and the crystal is no longer a single domain. However, ϵ_c must be a mean of $\epsilon_{\beta\beta}$ and $\epsilon_{\gamma\gamma}$, and ϵ_a must be a mean of all three dielectric constants. In the rhombohedral case we should expect

$$\epsilon_a = \epsilon_c = \frac{1}{3}(\epsilon_{\alpha\alpha} + 2\epsilon_{\gamma\gamma}).$$

However, Merz found ϵ_a to be much larger than ϵ_c . This may be the result of domain boundaries shifting under the application of a field. This would not affect the polarization in the c -direction since, presumably all domains have a positive component of polarization in that direction as

Fig. 5.

Principal dielectric constants of BaTiO₃ (theoretical).

the crystal had in the tetragonal region. On the other hand, the components in the α -direction would be randomly positive or negative, so a shift in boundary with fields might alter the net polarization and hence increase ϵ_a .

§ 8. PHENOMENOLOGICAL THEORY : HEAT OF POLARIZATION.

The polarization will be accompanied by a change of internal energy, which will be given by the usual formula

$$E = T \left(\frac{\partial A}{\partial T} \right)_p - A, \quad (8.1)$$

where A is given by (6.1). Now A depends on temperature partly through

the polarization, so we can write

$$\left(\frac{\partial A}{\partial T}\right)_p = \left(\frac{\partial A}{\partial T}\right)_{p,P} + \sum_{x,y,z} \left(\frac{\partial A}{\partial P_x}\right)_{p,T} \left(\frac{\partial P_x}{\partial T}\right)_p \quad (8.2)$$

$$= \left(\frac{\partial A}{\partial T}\right)_{p,P} + \sum E_x \left(\frac{\partial P_x}{\partial T}\right)_p, \quad (8.3)$$

and hence for zero field

$$\left(\frac{\partial A}{\partial T}\right)_p = \left(\frac{\partial A}{\partial T}\right)_{p,P} \quad (8.4)$$

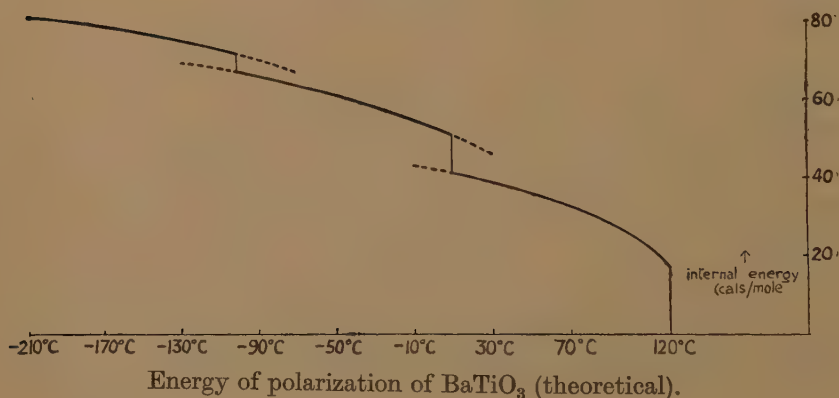
Now we have assumed all the coefficients to be independent of temperature except χ' , which is given by

$$\chi' = \chi'_0(T - T_0)/(T_1 - T_0). \quad (8.5)$$

Hence

$$T \left(\frac{\partial A}{\partial T}\right)_p = \frac{1}{2} \frac{\chi'_0 T}{T_1 - T_0} \{P_x^2 + P_y^2 + P_z^2\}. \quad (8.6)$$

Fig. 6.



We can now calculate the energy of polarization as a function of temperature using the values of the parameters we have already determined. The result is shown in fig. 6. Since the polarization is discontinuous at the transition temperatures there will be latent heats at these points. The calculated latent heats at the three transition temperatures, starting from the highest, are about 17 cal/mole, 10 cal/mole, and 5 cal/mole. These are lower than the experimental values. The latter, however, include some of the change in polarization energy which takes place below the transition temperatures.

§ 9. PHENOMENOLOGICAL THEORY: CRYSTAL FORM.

In order to study the strain which accompanies self polarization we must express the free energy as a function of polarization and strain. As before, we take the zero of free energy to be that of the unpolarized

cubic crystal. Then we have

$$\begin{aligned} A = & \frac{1}{2}c_{11}(x_x^2 + y_y^2 + z_z^2) + c_{12}(y_y z_z + z_z x_x + x_x y_y) + \frac{1}{2}c_{44}(x_y^2 + y_z^2 + z_x^2) \\ & + \frac{1}{2}\chi''(P_x^2 + P_y^2 + P_z^2) + \frac{1}{4}\xi_{11}''(P_x^4 + P_y^4 + P_z^4) + \frac{1}{2}\xi_{12}''(P_y^2 P_z^2 + P_z^2 P_x^2 + P_x^2 P_y^2) \\ & + g_{11}(x_x P_x^2 + y_y P_y^2 + z_z P_z^2) + g_{12}\{x_x(P_y^2 + P_z^2) + y_y(P_z^2 + P_x^2) + z_z(P_x^2 + P_y^2)\} \\ & + g_{44}(y_z P_y P_z + z_x P_z P_x + x_y P_x P_y). \quad \dots \dots \dots (9.1) \end{aligned}$$

All the terms of order P^4 or lower order are given, where it has been assumed that the strain is of the order of the square of the polarization. This will shortly be verified. Terms not given vanish because of the crystal symmetry.

The components of field and stress are given by the equations

$$X_x = \frac{\partial A}{\partial x_x}, \dots \quad E_x = \frac{\partial A}{\partial P_x}, \dots \quad \dots \quad (9.2)$$

If we assume that the six components of stress are zero, then we get six relations between strain and polarization, namely

$$\left. \begin{aligned} 0 = & c_{11}x_x + c_{12}(y_y + z_z) + g_{11}P_x^2 + g_{12}(P_y^2 + P_z^2), \\ & \dots \dots \dots \\ 0 = & c_{44}y_z + g_{44}P_y P_z, \\ & \dots \dots \dots \end{aligned} \right\} \dots \dots (9.3)$$

If we solve these equations for the strain in terms of the polarization we find that

$$\left. \begin{aligned} (c_{11} - c_{12})(c_{11} + 2c_{12})x_x = & P_x^2\{-g_{11}(c_{11} + c_{12}) + 2g_{12}c_{12}\} \\ & + (P_y^2 + P_z^2)(g_{11}c_{12} - g_{12}c_{11}), \\ & \dots \dots \dots \\ c_{44}y_z = & -g_{44}P_y P_z, \\ & \dots \dots \dots \end{aligned} \right\} \dots (9.4)$$

If we substitute these values for the strain in equation (9.1) then we get an expression for the free energy in terms of polarization for zero stress, that is, we obtain equation (6.1) except for terms of higher order than P^4 . Comparing the coefficients of P^2 etc., we find that

$$\left. \begin{aligned} \chi' = & \chi'', \\ \xi_{11}' = & \xi_{11}'' + 2 \frac{-g_{11}^2(c_{11} + c_{12}) + 4g_{11}g_{12}c_{12} - 2g_{12}^2c_{11}}{(c_{11} - c_{12})(c_{11} + 2c_{12})}, \\ \xi_{12}' = & \xi_{12}'' + 2 \frac{g_{11}^2c_{12} - 2g_{11}g_{12}c_{11} + g_{12}^2(-c_{11} + 2c_{12})}{(c_{11} - c_{12})(c_{11} + 2c_{12})} - \frac{g_{44}^2}{c_{44}}. \end{aligned} \right\} \dots (9.5)$$

From (9.4) we can see that the strains are proportional to the square of the polarization as already stated.

In the temperature ranges where the crystal is respectively cubic, tetragonal, orthorhombic and rhombohedral equations (9.4) become

$$\left. \begin{aligned} x_x = y_y = z_z = 0, \\ y_z = z_x = x_y = 0, \end{aligned} \right\} \quad (a)$$

$$\left. \begin{aligned} x_x = y_y = P_z^2 \frac{g_{11}c_{12} - g_{12}c_{11}}{(c_{11} - c_{12})(c_{11} + 2c_{12})}, \\ z_z = P_z^2 \frac{-g_{11}(c_{11} + c_{12}) + 2g_{12}c_{12}}{(c_{11} - c_{12})(c_{11} + 2c_{12})}, \\ y_z = z_x = x_y = 0, \end{aligned} \right\} \quad (b)$$

$$\left. \begin{aligned} x_x = P_z^2 \frac{2(g_{11}c_{12} - g_{12}c_{11})}{(c_{11} - c_{12})(c_{11} + 2c_{12})} \\ y_y = z_z = P_z^2 \frac{-g_{11}c_{11} - g_{12}(c_{11} - 2c_{12})}{(c_{11} - c_{12})(c_{11} + 2c_{12})} \\ z_x = x_y = 0, \quad y_z = -g_{44}^2 P_z^2 / c_{44}, \end{aligned} \right\} \quad (c)$$

$$\left. \begin{aligned} x_x = y_y = z_z = P_z^2 \frac{-g_{11} - 2g_{12}}{c_{11} + 2c_{12}}, \\ y_z = z_x = x_y = -g_{44} P_z^2 / c_{44}. \end{aligned} \right\} \quad (d)$$

$$\left. \begin{aligned} (a) \\ (b) \\ (c) \\ (d) \end{aligned} \right\} \quad \dots \quad (9.6)$$

From the observed strains and polarizations we can get some information about the constants. The most convenient quantity to use is the difference between the strains along the c and a -axes, that is $z_z - x_x$, which is given by

$$z_z - x_x = P_z^2 (g_{12} - g_{11}) / (c_{11} - c_{12}), \quad \dots \quad (9.7)$$

in either the tetragonal or the orthorhombic region. This is about 0.01 at room temperature when P_z is about 16 microcoulombs. Hence we have

$$g_{12} - g_{11} = (c_{11} - c_{12}) \times 4 \times 10^{-12}. \quad \dots \quad (9.8)$$

Similarly in the tetragonal region we have for the volume expansion

$$x_x + y_y + z_z = P_z^2 (-g_{11} - 2g_{12}) / (c_{11} + 2c_{12}). \quad \dots \quad (9.9)$$

The volume expansion is more difficult to determine since we have to estimate the unstrained volume by extrapolation from the cubic region. However, we have approximately

$$-(g_{11} + 2g_{12}) = (c_{11} + 2c_{12}) \times 0.8 \times 10^{-12}. \quad \dots \quad (9.10)$$

The shear y_z just below the second transition temperature is about 14'. Hence we have

$$-g_{44} = c_{44} \times 2.7 \times 10^{-12}. \quad \dots \quad (9.11)$$

In the orthorhombic region we have

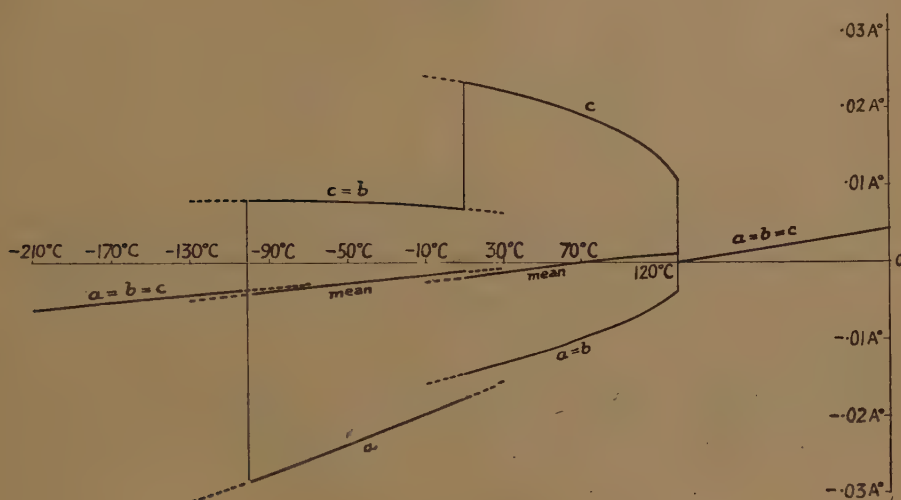
$$x_x + y_y + z_z = 2P_z^2 (-g_{11} - 2g_{12}) / (c_{11} + 2c_{12}). \quad \dots \quad (9.12)$$

Hence in all regions the strains can be evaluated in terms of P_z^2 and the ratios $(g_{12} - g_{11}) / (c_{11} - c_{12})$ and $(g_{11} + 2g_{12}) / (c_{11} + 2c_{12})$ given empirically

by (9.8) and (9.10). The calculated strains are given in fig. 7. A term representing the thermal expansion is added so that the figure shows the total variation of axial length with temperature.

By considering the observed piezoelectric resonances Mason (1948) has been able to calculate the g 's and c 's independently for the ceramic. These quantities might, however, be very different for the single crystal, so we are not able to make use of his results.

Fig. 7.



Lattice spacing of BaTiO_3 relative to spacing at 120°C . (theoretical).

§ 10. MODEL THEORY.

Our picture of barium titanate is that it is an ionic crystal, and we shall follow Born's treatment of such crystals as given by Fowler (1936). The ions are regarded as point centres of force, and the potential energy of two ions with charges ϵ_1 and ϵ_2 at a distance r apart is assumed to have the form

$$\epsilon_1 \epsilon_2 r^{-1} - \mu r^{-6} + \lambda r^{-9}.$$

The second term represents the Van der Waals attraction, and the third term the repulsive forces. The potential energy of BaTiO_3 per unit cell is then given by the expression

$$\begin{aligned} \phi(R) = & \frac{-49 \cdot 1 \epsilon^2}{R} + \frac{3 \cdot 31 (\lambda_{\text{TT}} + \lambda_{\text{BB}} + 3\lambda_{\text{OO}})}{R^9} + \frac{8 \cdot 07 \lambda_{\text{BT}}}{(2R/\sqrt{3})^9} + \frac{6 \cdot 07 \lambda_{\text{OT}}}{(R/2)^9} \\ & + \frac{12 \cdot 30 (\lambda_{\text{OB}} + \lambda_{\text{OO}})}{(R/\sqrt{2})^9} - \frac{4 \cdot 20 (\mu_{\text{TT}} + \mu_{\text{BB}} + 3\mu_{\text{OO}})}{R^6} - \frac{8 \cdot 71 \mu_{\text{BT}}}{(2R/\sqrt{3})^6} \\ & - \frac{6 \cdot 24 \mu_{\text{OT}}}{(R/2)^6} - \frac{13 \cdot 29 (\mu_{\text{OB}} + \mu_{\text{OO}})}{(R/\sqrt{2})^6}, \quad \dots \dots \dots (10.1) \end{aligned}$$

where we have assumed the unit cell to be cubic with an edge of length R ,

and the charges on the Ba, Ti and O ions to be 2ϵ , 4ϵ and -2ϵ respectively, $-\epsilon$ being the electronic charge. The values of λ_{BB} , λ_{OO} , λ_{OB} , μ_{BB} , μ_{OB} , and μ_{OO} have been given by Fowler (1936). We have used his methods to estimate λ_{OT} and μ_{OT} , and the contributions of the terms containing λ_{TT} , λ_{BT} , μ_{TT} and μ_{BT} are negligible. The complete set of values used is given in Table I.

TABLE I.

Force constants between ions in c.g.s. units.

λ_{BB}	82.7×10^{-82}	μ_{BB}	239.0×10^{-60}
λ_{OB}	99.0×10^{-82}	μ_{OB}	162.0×10^{-60}
λ_{OO}	113.5×10^{-82}	μ_{OO}	135.0×10^{-60}
λ_{OT}	15.6×10^{-82}	μ_{OT}	31.3×10^{-60}

If we choose R so that $\phi(R)$ is a minimum this will give us the side of the unit cell at zero temperature. Using the force constants from Table I. we find that R is 4.04 \AA . as compared with the extrapolated observed value of 3.99 \AA . This suggests that our force constants are reasonably accurate, and we can now use equation (10.1) to estimate the compressibility.

The change in potential energy of the unit cell when we alter R by a small amount ΔR is $\frac{1}{2}(\Delta R)^2 \phi''(R)$, and hence the change per unit volume is $\frac{1}{2}(\Delta R)^2 \phi''(R) R^{-3}$. But the change in free energy is

$$\frac{1}{2} c_{11} (x_x^2 + y_y^2 + z_z^2) + c_{12} (x_x y_y + y_y z_z + z_z x_x),$$

where

$$x_x = y_y = z_z = \Delta R / R.$$

If we neglect the difference between free and potential energy these two quantities can be equated, and we have

$$c_{11} + 2c_{12} = \frac{1}{3} R^{-1} \phi''(R), \quad . \quad . \quad . \quad . \quad . \quad (10.2)$$

Putting in the values for the force constants from Table I. we have

$$c_{11} + 2c_{12} = 1.2 \times 10 \text{ dynes/cm}^2. \quad . \quad . \quad . \quad . \quad (10.3)$$

As our assumption that the ions are point centres of force can only be a rough approximation to the truth we shall not attempt to calculate c_{11} and c_{12} separately but shall assume c_{11} to be equal to $2c_{12}$. This is equivalent to be taking Poisson's ratio to be $\frac{1}{3}$, which is not far from the value for most substances. This makes

$$\left. \begin{aligned} c_{11} &= 6.0 \times 10^{12} \text{ dynes/cm}^2, \\ c_{12} &= 3.0 \times 10^{12} \text{ dynes/cm}^2, \end{aligned} \right\} . \quad . \quad . \quad . \quad . \quad (10.4)$$

and hence from (10.8) and (10.10),

$$\left. \begin{aligned} g_{11} + 2g_{12} &= -9.6, \\ g_{11} &= -11.2, \\ g_{12} &= -0.8. \end{aligned} \right\} . \quad . \quad . \quad . \quad . \quad (10.5)$$

If we assume central forces between the ions c_{44} should have the same value as c_{12} . This makes

$$g_{44} = -8.$$

Substituting the above values in equations (9.5) we get

$$\xi'_{11} = \xi''_{11} - 70 \times 10^{-12},$$

and

$$\begin{aligned} \xi'_{12} &= \xi''_{12} + 26.8 \times 10^{-12} - 21.3 \times 10^{-12} \\ &= \xi''_{12} + 5.5 \times 10^{-12}. \end{aligned}$$

Putting in the values of ξ'_1 and ξ'_{12} from (7.14) and (7.15) this gives

$$\left. \begin{aligned} \xi''_{11} &= -4.4 \times 10^{-12} + 70 \times 10^{-12} = 65.6 \times 10^{-12}, \\ \xi''_{12} &= 5.3 \times 10^{-12} - 5.5 \times 10^{-12} = -0.2 \times 10^{-12}. \end{aligned} \right\} \quad (10.6)$$

We shall now examine the field in which each ion moves. We assume all the ions except one to be fixed, and then it is a straightforward matter to calculate the field in which that ion moves in terms of the force constants. If the ion is displaced a distance r from its symmetrical position then its potential energy can be expanded in the form

$$r^2 a(\theta, \phi) + r^4 b(\theta, \phi),$$

where θ and ϕ are polar coordinates referred to some suitable axis. For barium and titanium, which are symmetrically placed, a is independent of θ and ϕ and is given by

$$\begin{aligned} a_B &= 1.2 \times 10^5 \text{ ergs/cm.}^2, \\ a_T &= 4.7 \times 10^5 \text{ ergs/cm.}^2. \end{aligned}$$

The contribution to a_B and a_T from the short range repulsive forces, which is positive, is several times larger than the contribution from the electrostatic and Van der Waal's forces. Thus a_B and a_T are bound to be positive. The case is different for the oxygen ion. This lies in a much less symmetrical field, and if we take the initial line to be the line joining the O ion to a neighbouring Ti ion we find that

$$a_O = 0.6 \times 10^5 [(7.4 \cos^2 \theta + 4.2 \sin^2 \theta) - \alpha R^{-3} 196.4 (\cos^2 \theta + \frac{1}{4} \sin^2 \theta)] \text{ ergs/cm.}^2, \quad (10.7)$$

where α is the polarizability of the oxygen ion, and R is the edge of the unit cell. The term in α is there because if the ion is displaced there will be an electrostatic field acting on it proportional to its displacement. This will polarize the ion and there will be a lowering of potential energy proportional to the square of the displacement. There is no similar term for the barium or titanium ions because, owing to the symmetrical arrangement of the ions round them, the electrostatic field acting on a displaced ion is proportional to the cube of the displacement. Various values have been given for α . A value of about 2.4×10^{-24} gives the correct value for the refractive index of CaTiO_3 , after allowing for the polarizability of the Ca and Ti ions. This value makes a_O about zero along $\theta=0$, and positive for other directions. The value of α given above, however, is only a mean value and is likely to be different for different

directions since the oxygen ion must be very distorted. Hence we cannot predict with certainty whether a_0 will be always positive or may have negative values in certain directions. In the latter case, of course, the symmetrical position of the oxygen ion will be unstable, and there will be positions of equilibrium off centre. These are likely to be in the direction $\theta=0$, that is, in the direction of the neighbouring Ti ions.

We must now consider the relation between the constants a , b , etc., and the free energy. This involves studying the thermal vibrations of the substance. We assume that the vibrations of each ion are independent. Let

$$a_x x^2 + a_y y^2 + a_z z^2 + b_{xx} x^4 + b_{yy} y^4 + b_{zz} z^4 + 2b_{yz} y^2 z^2 + 2b_{zx} z^2 x^2 + 2b_{xy} x^2 y^2$$

be the potential energy of an ion when displaced to a point (x, y, z) relative to the symmetrical position as origin, the substance being unpolarized. The above expression includes all terms of lower order than the sixth. The displaced ion is equivalent to a dipole which we assume to have components $(\gamma_x x, \gamma_y y, \gamma_z z)$ where the γ 's are constants of the order of the ionic charge. Now if the substance is polarized there will be a force acting on the ion having components $(\beta P_x, \beta P_y, \beta P_z)$, where β is the Lorentz factor, and hence an additional term in the potential energy of magnitude

$$-\beta(\gamma_x x P_x + \gamma_y y P_y + \gamma_z z P_z).$$

Then the partition function f for a single ion is given by

$$f = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \exp \{ [\beta \gamma_x x P_x + \dots - a_x x^2 - \dots - b_{xx} x^4 - \dots - 2b_{yz} y^2 z^2 - \dots] / kT \} dx dy dz. \quad (10.8)$$

We shall assume that $b kT/a^2$ is small. Then terms in b^2 can be neglected and we have

$$f = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left(1 - \frac{b_{xx} x^4 \dots + 2b_{yz} y^2 z^2 + \dots}{kT} \right) \exp \left[\frac{\beta \gamma_x x P_x + \dots - a_x x^2 - \dots}{kT} \right] dx dy dz. \quad (10.9)$$

If we now put

$$x = \frac{1}{2} \beta \gamma_x P_x a_x^{-1} + x', \quad \dots \dots \dots (10.10)$$

then

$$\begin{aligned} f = & \exp \left\{ \frac{1}{4} \frac{\beta^2 \gamma_x^2 P_x^2}{a_x kT} + \dots \right\} \\ & \times \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \left\{ 1 - \frac{b_{xx}}{kT} \left(\frac{\beta^4 \gamma_x^4 P_x^4}{16 a_x^4} + \frac{3}{2} \frac{\beta^2 \gamma_x^2 P_x^2 x^2}{a_x^2} + x^4 \right) - \dots \right. \\ & \left. - \frac{2b_{yz}}{kT} \left(\frac{\beta^4 \gamma_y^2 \gamma_z^2 P_y^2 P_z^2}{16 a_y^2 a_z^2} + \frac{\beta^2 \gamma_y^2 P_y^2 z^2}{4 a_y^2} + \frac{\beta^2 \gamma_z^2 P_z^2 y^2}{4 a_z^2} + y^2 z^2 \right) \right\} \\ & \times \exp \left\{ \frac{-a_x x^2 - \dots}{kT} \right\} dx dy dz \end{aligned}$$

$$\begin{aligned}
&= \frac{(\pi kT)^{3/2}}{(a_x a_y a_z)^{1/2}} \exp \left\{ \frac{1}{4} \frac{\beta^2 \gamma_x^2 P_x^2}{a_x kT} + \dots \right\} \\
&\times \left\{ 1 - \frac{3}{4} \frac{b_{xx} kT}{a_x^2} - \dots - \frac{1}{2} \frac{b_{yz} kT}{a_y a_z} - \dots - \frac{3}{4} \frac{b_{xx} \beta^2 \gamma_x^2 P_x^2}{a_x^3} - \dots \right. \\
&- \frac{1}{4} b_{yz} \left(\frac{\beta^2 \gamma_y^2 P_y^2}{a_y^2} + \frac{\beta^2 \gamma_z^2 P_z^2}{a_z^2} \right) - \dots - \frac{b_{xx}}{kT} \frac{\beta^4 \gamma_x^4 P_x^4}{16 a_x^4} - \dots \\
&\left. - \frac{b_{yz}}{kT} \frac{\beta^4 \gamma_y^2 \gamma_z^2 P_y^2 P_z^2}{8 a_y^2 a_z^2} - \dots \right\}. \quad (10.11)
\end{aligned}$$

The free energy per unit volume is given by

$$A = -NkT \Sigma \log f + \frac{1}{2} \beta (P_x^2 + P_y^2 + P_z^2), \quad (10.12)$$

where N is the number of unit cells per unit volume, and the summation is over the different ions in the unit cell. The term in P^2 arises because in summing over the unit cells we have reckoned twice over the equilibrium potential energy of the ions, which is $-\frac{1}{2} \beta (P_x^2 + P_y^2 + P_z^2)$. Hence we have

$$\begin{aligned}
A = N \Sigma \left[-\frac{1}{2} kT \log \frac{(\pi kT)^3}{a_x a_y a_z} + (kT)^2 \left(\frac{3}{4} \frac{b_{xx}}{a_x^2} + \dots + \frac{1}{2} \frac{b_{yz}}{a_y a_z} + \dots \right) \right] \\
+ P_x^2 \left[\frac{1}{2} \beta - \frac{1}{4} N \beta^2 \Sigma \left(\frac{\gamma_x^2}{a_x} - 3kT \frac{b_{xx} \gamma_x^2}{a_x^3} - kT \frac{b_{xy} \gamma_x^2}{a_x a_y} - kT \frac{b_{xz} \gamma_x^2}{a_x^2 a_z} \right) \right] + \dots \\
+ P_x^4 N \beta^4 \Sigma \frac{b_{xx} \gamma_x^4}{a_x^4} + \dots + P_y^2 P_z^2 N \beta^4 \Sigma \frac{b_{yz} \gamma_y^2 \gamma_z^2}{a_y^2 a_z^2} + \dots \quad (10.13)
\end{aligned}$$

If we compare this expression with equation (9.1) we have

$$\chi'' = \beta - \frac{1}{2} N \beta^2 \Sigma \left\{ \frac{\gamma_x^2}{a_x} - \frac{kT \gamma_x^2}{a_x^2} \left(3 \frac{b_{xx}}{a_x} + \frac{b_{xy}}{a_y} + \frac{b_{xz}}{a_z} \right) \right\}, \quad (10.14)$$

$$\xi'_{11} = \frac{N \beta^4}{4} \Sigma \frac{b_{xx} \gamma_x^4}{a_x^4}, \quad (10.15)$$

$$\xi'_{12} = \frac{N \beta^4}{4} \Sigma \frac{b_{xy} \gamma_y^2 \gamma_z^2}{a_z^4}. \quad (10.16)$$

As before, the summation is over the different ions in the unit cell, and when this summation is carried out the quantities involved will be the same for all axes (or pairs of axes) owing to the crystal symmetry. It is possible to make rough estimates of the b 's and a 's in terms of the force constants in the way already described, and we find that the term involving T in (10.14) is small compared with the other terms at constant temperatures. Now we have already found that χ'' has to vanish at a temperature T_0 , so that approximately

$$1 = \frac{1}{2} N \beta \Sigma \frac{\gamma_x^2}{a_x}. \quad (10.17)$$

Now from (9.1) we see that the coefficients g_{11} and g_{12} measure the variation of $\frac{1}{2} \chi''$ with strain. Now as we have already stated the third

term in χ'' is small and β , being the Lorentz factor, will not depend much on strain, and hence we have approximately

$$g_{11} = \frac{1}{4} N \beta^2 \Sigma \frac{\gamma_x^2}{a_x} \frac{\partial}{\partial x_x} \log \left(\frac{a_x}{\gamma_x^2} \right), \quad (10.18)$$

and
$$g_{12} = \frac{1}{4} N \beta^2 \Sigma \frac{\gamma_x^2}{a_x} \frac{\partial}{\partial y_y} \log \left(\frac{a_x}{\gamma_x^2} \right). \quad (10.19)$$

Hence
$$g_{11} + 2g_{12} = \frac{1}{4} N \beta^2 \Sigma \frac{\gamma_x^2}{a_x} \left(\frac{\partial}{\partial x_x} + \frac{\partial}{\partial y_y} + \frac{\partial}{\partial z_z} \right) \log \left(\frac{a_x}{\gamma_x^2} \right), \quad . (10.20)$$

where R is the side of the unit cell. Now a_x decreases as about the tenth power of R , and γ_x will not vary much with R . Hence from (10.20) and (10.17) we have

$$g_{11} + 2g_{12} \approx -5\beta \approx -20, \quad (10.21)$$

if we take β , the Lorentz factor, to have the value $4\pi/3$. This differs from the value given in (10.5) by a factor of 2. In view of the approximate nature of both calculations this is however, not surprising. From (10.19) we see that the small value of g_{12} given in (10.5) means that a_x is little changed by a strain in the y -direction. This is what we should expect. Again we found that χ'' was given by an equation of the form

$$\chi'' = \chi' = \chi'_0 (T - T_0) / (T_1 - T_0),$$

under conditions of zero stress, that is $\chi'_0 / (T_1 - T_0)$ is the rate of variation of χ'' with temperature, when the stress is zero. Hence

$$\begin{aligned} \frac{\chi'_0}{T_1 - T_0} &= \frac{\partial \chi''}{\partial T} + g_{11} \frac{dx_x}{dT} + g_{12} \left(\frac{dy_y}{dT} + \frac{dz_z}{dT} \right) \\ &= \frac{\partial \chi''}{\partial T} + \alpha (g_{11} + 2g_{12}), \quad (10.22) \end{aligned}$$

where α is the coefficient of thermal expansion and the partial differentiation with respect to T is for constant strain. This follows because g_{11} and g_{12} give the variation with strain. Hence from (11.14) it follows that

$$\frac{\chi'_0}{T_1 - T_0} = \frac{1}{2} N \beta^2 k \Sigma \frac{\gamma_x^2}{a_x^2} \left(\frac{3b_{xx}}{a_x} + \frac{b_{xy}}{a_y} + \frac{b_{xz}}{a_z} \right) + \alpha (g_{11} + 2g_{12}). \quad (10.23)$$

The thermal expansion coefficient α can be obtained from the X-ray data above the highest transition temperature and is about 10^{-5} . Hence, since $g_{11} + 2g_{12}$ is approximately -10 , the value of $\alpha (g_{11} + 2g_{12})$ is approximately -10^{-4} , and since the left-hand side is 10^{-4} the first term on the right-hand side must be 2×10^{-4} . It can be verified that it is of this order. Physically the whole expression in (10.23) gives the rate of decrease of polarizability with temperature. The second term gives the indirect effect of temperature change through the medium of lattice expansion. It corresponds to an increase in polarizability, as is physically obvious since lattice expansion must allow the ions to move more easily. The first term gives the direct effect of thermal vibrations. We see that they cause the polarizability to decrease, and this term is comparable with, though larger than, the other.

This work is part of a project carried out with the financial assistance of the Electrical Research Association, and is published with their permission.

REFERENCES.

- BLATTNER, KANZIG, and MERZ, 1949, *Helv. Phys. Acta*, **22**, 35.
 CADY, *Piezoelectricity* (New York: McGraw-Hill & Co.).
 FOWLER, 1936, *Statistical Mechanics* (2nd ed.), (Oxford: University Press).
 GINSBERG, 1946, *J. Phys. U.S.S.R.*, **10**, 107.
 HARWOOD, POPPER, and RUSHMAN, 1947, *Nature*, **160**, 59.
 HULM, 1947, *Nature*, **160**, 127.
 KAY, 1948, *Acta Cryst.*, **1**, 229.
 KAY, VOUSDEN, and WELLARD, 1949, *Nature*, **163**, 637.
 JACKSON, and REDDISH, 1945, *Nature*, **156**, 717.
 MASON, 1948, *Phys. Rev.*, **74**, 1134.
 MASON, and MATTHIAS, 1948, *Phys. Rev.*, **74**, 1622.
 MATTHIAS, and VON HIPPEL, 1948, *Phys. Rev.*, **73**, 1378.
 MEGAW, 1946, *Trans. Faraday Soc.*, **42A**, 224.
 MERZ, 1949, *Phys. Rev.*, **75**, 687.
 MUELLER, 1940, *Phys. Rev.*, **57**, 829; **58**, 565, 805.
 RUSHMAN, and STRIVENS, 1946, *Trans. Faraday Soc.*, **42A**, 231.
 VON HIPPEL, BROCKENRIDGE, CHELSEY, and TISZA, 1946, *Ind. & Eng. Chem.*, **38**, 1097.
 WÜL, 1946, *J. Phys. U.S.S.R.*, **10**, 95.

XCVII. Theory of the High Pressure Helium Discharge.

By V. J. FRANCIS, B.Sc., F.Inst.P., M.I.E.E., A.R.C.S.

(Communication from the Staff of the Research Laboratories of The General Electric Company Limited, Wembley, England.)

[Received July 11, 1949.]

SUMMARY.

The method developed in recent papers for calculating the characteristics of the high pressure discharge is extended to helium. The properties of the helium discharge derived in this way are compared with those of high pressure mercury vapour, and it is shown that a loading of the order of 10^4 watts per cm. of arc is required in the former to produce conditions usually associated with the normal operation of the high pressure mercury vapour discharge at 10 to 100 watts per cm. of arc.

§1. INTRODUCTION.

IN some recent papers a method has been developed for calculating approximately the properties of the high pressure discharge (Francis 1946 a, b, 1949). The nomenclature and symbols employed in these publications are used also in the present paper. The method is approximate only since it is assumed (*a*) that all the energy is evolved within a certain radius r_1 of the discharge column and that over that area it is uniformly distributed. The radius r_1 , is, of course, a function of the operating conditions of the discharge. It is assumed also that (*b*) heat losses from

the arc other than by radiation can be accounted for by a coefficient of conduction of the form

$$\sigma(T) = \sigma_0 T, \quad (1.1)$$

and that (c) radiation losses are given by a term depending upon the value of a fictitious excitation level, the value of this excitation level as well as the ionization potential remaining constant whatever the conditions of operation.

In addition to the inaccuracies of a mathematical nature introduced by the assumption (a), the approximations (b) and (c) make it impossible that characteristics so calculated can be accurate even if, as is not the case, the atomic parameters were known with certainty. It is possible, however, by comparing calculated characteristics at one or two points with measured characteristics, to verify the value of, or even to deduce, the required constants, and in this way the calculated characteristics may be extended with reasonable confidence over a wide range and into regions where experiments cannot easily be made.

The calculations so far have been confined to mercury (Francis 1946 a, b) and cadmium (Francis 1949). There are, however, other gases and vapours on which data of the kind which can be obtained by these calculations would be useful—particularly in circumstances in which it is difficult or impossible to obtain experimental results. There is practical interest for example, in the discharge through the rare gases at high pressure (Aldington), and as part of a programme on these, the results for helium given in this paper were obtained.

§2. METHOD OF CALCULATION.

The calculations follow the same lines as those for mercury and cadmium (Francis 1949). There are, however, a few differences. In the earlier papers the excitation function $\log_{10} g_e g^+ / g$ was omitted from the Saha equation. Here $g = g_0 + \sum_n g_n [\exp -(\epsilon_n - \epsilon_0)/kt]$ where g_n is the weight $(2j+1)$ of the excited state of energy ϵ_n and g_0, ϵ_0 refer to the ground state. The quantities g_e and g^+ are the similar sums for the electron and the positive ion respectively. This omission made no serious error to the results in the earlier papers since there was always a disposable constant whose value had to be determined by experiment, and in the comparison of cadmium with mercury (Francis 1949) the term had the same value for each. The only material effect on the results was an incorrect labelling of the $Er_0^{3/2}$ parameter in the EI/T_0 curves in fig. 9 of reference (Francis 1949), the values of which should be reduced by about 25 per cent, so that they range from 6 to 30 instead of from 8 to 40. The term is included here, however, as use will later be made of it. The Saha equation becomes

$$x = 10^{-\left(\frac{2.53 \times 10^8 V_i}{T} + 3.25\right)} G^{1/2} p^{1/2} T^{5/4}. \quad (2.1)$$

Where $G = g_e g^+ / g$ and the $G^{1/2}$ factor was that omitted in the earlier papers.

The other variables in (2.1) are x , V_i , T and p , the degree of ionization, the ionization potential, the temperature and the pressure respectively. In the calculations, G is retained as one of the constants characterizing the gas. The ionization potential (V_i) of helium is 24.47 V. The energy of the mean fictitious level (V_m) was taken as 22 V. The results for helium are therefore of interest both because the ionization potential is so much higher than that of mercury and cadmium, and also because the excitation levels are relatively little below the ionization potential. This latter fact will, incidentally, probably result in a relatively larger contribution of the recombination continuous spectrum to the radiation term, and may therefore lead to larger variations in the effective value of V_m at high loadings than might be expected for mercury or cadmium. The fact that V_m is assumed constant means therefore that, at high loadings, larger inaccuracies are to be expected than for mercury and cadmium. From the values of V_i and V_m , the functions $\lambda(T)$ and $\phi(T)$ could be calculated. To obtain $\mu(T)$ and μ_0 it was first of all necessary to obtain an estimate of the range of $A_3 P^{3/2}/G^{1/2} E^2$ and β likely to be required; E is here the arc potential gradient and β is given by $\beta = K_1 G^{1/2} E^2 P^{-1/2} \mu_0 r_1^2/2$. This was done by calculating from estimated values of the atomic parameters, and from the range of P and E required, the likely values of $A_3 P^{3/2}/G^{1/2} E^2$. A few points were then worked out, similar to those in Table I. of reference (Francis 1946 b), to find the region in which occurred the maximum value of μ_0 , as it was this region for the helium discharge, corresponding to the similar region for the normal operating conditions of the mercury vapour discharge, about which information was required. At these values of β a few values of A_3 , K_1 $P r_0^2$ were found in order to verify that the values of the pressure were in the required range. It was then a simple matter to extend the calculations to the values of $A_3 P^{3/2}/G^{1/2} E^2$ and β to cover the interesting range of operating conditions. The values of $A_3 P^{3/2}/G^{1/2} E^2 \times 10^{-8}$ finally adopted were 20, 60 and 100, and three sets were calculated for values of β ranging from about 50 to 600 at intervals of 50.

§ 3. CHARACTERISTICS SHOWING THE EFFECT OF THE IONIZATION POTENTIAL AND EXCITATION LEVEL.

From the sets of tables similar to Table I. of reference (Francis 1946 b), elimination was carried out in the same way as previously for mercury and cadmium, the constants being left in the derived expressions. The opportunity was taken however to change the constants A_3 , K_1 , into those more closely related to the actual atomic parameters by means of the relations

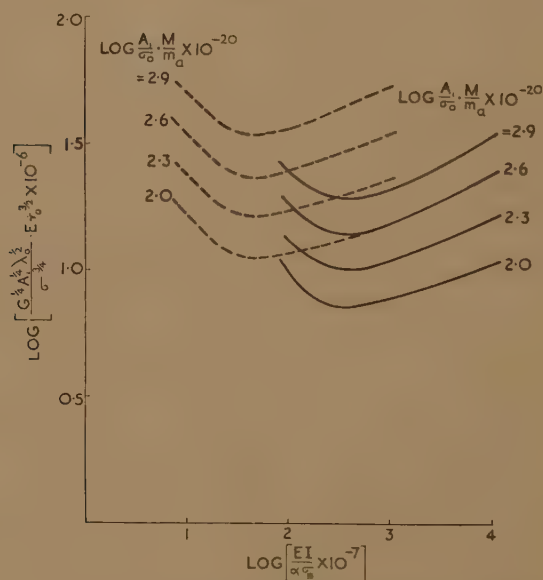
$$\left. \begin{aligned} K_1 &= 0.0109 \lambda_0 / \sigma_0, \\ A_3 &= 1.33 \times 10^{11} A_1 / \lambda_0. \end{aligned} \right\} \quad \dots \dots \dots (3.1)$$

The results are shown in figs. 1, 2 and 3, where the full lines are for the helium discharge, and the dotted lines give the characteristics for mercury vapour for comparison. Since all the constants which differentiate helium from mercury with the exception of the excitation and ionization

potentials remain in the expressions, the differences shown in these four figures are due entirely to the change in these two characteristic potentials.

Fig. 1 gives the E/EI characteristic at constant density. The effect of the greater ionization potential for helium is to move the minima of the curves to larger values of $EI/\alpha\sigma_0$. This would be expected, since, in general terms, the "positive" part of the characteristic is accounted for by the rapidly increasing proportion of the input energy that is lost as radiation at the higher temperatures, leaving relatively less energy available to heat and ionize the gas. This stage would not be reached in helium until larger values of $EI/\alpha\sigma_0$ than those required in mercury. The value of the $Er_0^{3/2}$ function for given value of $A_1/\sigma_0 \cdot M/ma$ at the minimum is lower for helium than for mercury. This is probably due to the fact

Fig. 1.



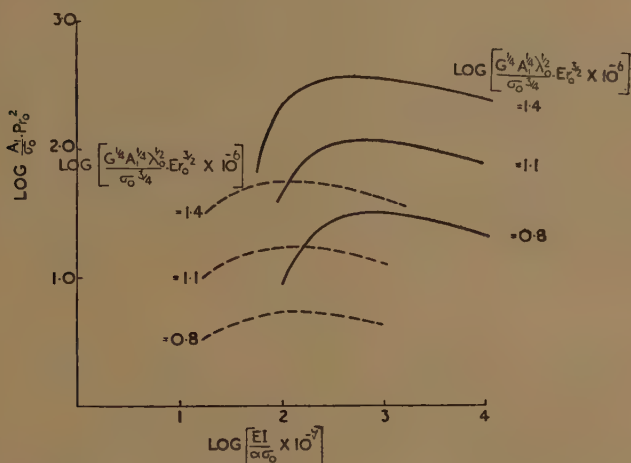
Arc voltage—arc loading (watts) characteristic at constant density ;
 ————— helium.
 - - - - - mercury vapour.

that the excitation potential for helium lies much closer to the ionization potential than in the case of mercury. The result is that, at the minimum, when the proportion of radiated energy to conducted energy is similar for the two gases, the degree of ionization is higher for helium than for mercury.

Fig. 2 gives the EI/P characteristic at constant E . Again the turnover point occurs at larger values of $EI/\alpha\sigma_0$ for helium than for mercury, and for the same reason as that associated with the change of the minima in fig. 1. The increase of $A_1/\sigma_0 \cdot Pr_0^2$ at the turnover point for a given value of the $Er_0^{3/2}$ function is a little more difficult to understand, but it is almost certainly the result of the much higher temperature required to produce

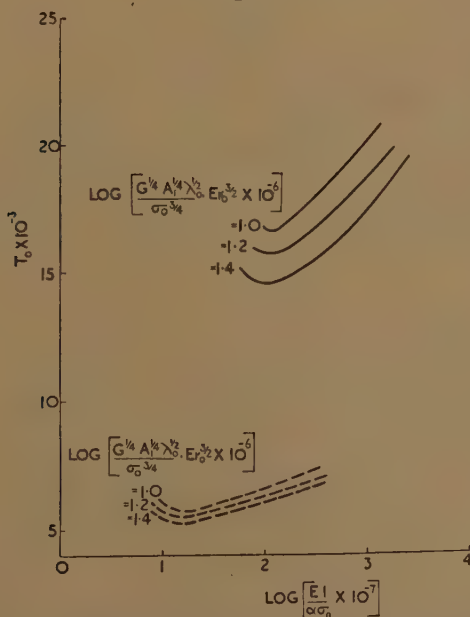
the necessary ionization in helium. The interpretation of the curves in fig. 3 giving the EI/T_0 characteristics is quite obvious ; the minima moved

Fig. 2.



Pressure—arc loading (watts) characteristic at constant arc voltage ;
 ————— helium.
 - - - - - mercury vapour.

Fig. 3.



Temperature—arc loading (watts) characteristic at constant arc voltage ;
 ————— helium.
 - - - - - mercury vapour.

to higher values of $EI/\alpha\sigma_0$ for the reasons already given, and the larger values of T_0 being a direct consequence of the higher ionization potential.

§4. THE CONSTANTS FOR HELIUM.

To obtain absolute values of the characteristics, it is necessary to know the value of the constants occurring in the expressions in figs 1 to 3. The method adopted here was to compare the best obtainable values of the constants for helium with those for mercury, and since the values of the latter are known with some accuracy, an estimate of the former may be made. It is not suggested that there is any certainty that the values so obtained are closely characteristic of helium, but it will be seen that an experimental check gives some degree of confidence.

If we assume that the " f "-values for mercury and for helium are the same—and such measurements as have been made do not suggest that this assumption is seriously incorrect—we have

$$A_1/A'_1 = \nu^3/\nu'^3 = (22/8)^3, \quad (4.1)$$

where the undashed quantities refer to helium and the dashed quantities to mercury. The ν 's refer, of course, to the wave number of the transition from the fictitious energy levels to the ground states.

So far as the mean free paths are concerned, the cross section for mercury (Brode 1930) at 6000°K is about $400 \text{ cm.}^2/\text{cm.}^3$ and that for helium (Kollath 1930) at $16,000^\circ \text{K}$ is about $38 \text{ cm.}^2/\text{cm.}^3$. We are probably therefore, not very far out if we write

$$\lambda_0 = 10\lambda'_0. \quad (4.2)$$

For the thermal conductivity, there is not much difficulty. Values for mercury and helium at ordinary temperatures are available, 1.85×10^5 at 203°C . for the former, and 34.3×10^5 at 0°C . for the latter, the units in each case being cal. per sec. per $^\circ \text{C}$. Whatever the temperature function (1.1) really should be, it is not likely to differ much for the two different gases. It is therefore found that

$$\sigma_0 = 32.3\sigma'_0. \quad (4.3)$$

Although the introduction of the G factor makes little alteration to the characteristics for mercury and cadmium found in the earlier papers, it does, of course, alter somewhat the value of the constants, and for mercury are found:

$$\left. \begin{aligned} A_1/\sigma_0 &= 6.46 \times 10^3, \\ G^{1/4} A_1^{1/4} \lambda_0^{1/2} / \sigma_0^{3/4} &= 3.34 \times 10^{-2}, \\ \sigma_0 &= 0.95, \end{aligned} \right\} (4.4)$$

and it follows that, for helium:

$$\left. \begin{aligned} A_1/\sigma_0 &= 4.15 \times 10^3, \\ G^{1/4} A_1^{1/4} \lambda_0^{1/2} / \sigma_0^{3/4} &= 1.65 \times 10^{-2}, \\ \sigma_0 &= 30.75. \end{aligned} \right\} (4.5)$$

§ 5. EXPERIMENTAL CHECK ON VALUE OF HELIUM CONSTANTS.

It was not thought worth while to proceed further without experimental verification that the constants deduced above were reasonable. There was difficulty in obtaining this experimental verification. In the first place, the high pressure regime in which the interest lies makes it necessary to work at helium pressures not much less than one atmosphere. It would in fact be better to work at higher pressures but, in addition to the experimental inconvenience of working much above one atmosphere, this would only increase the second difficulty, namely that the loadings required are so high, that even with silica envelopes it is impossible to use other than spot readings.

TABLE I.

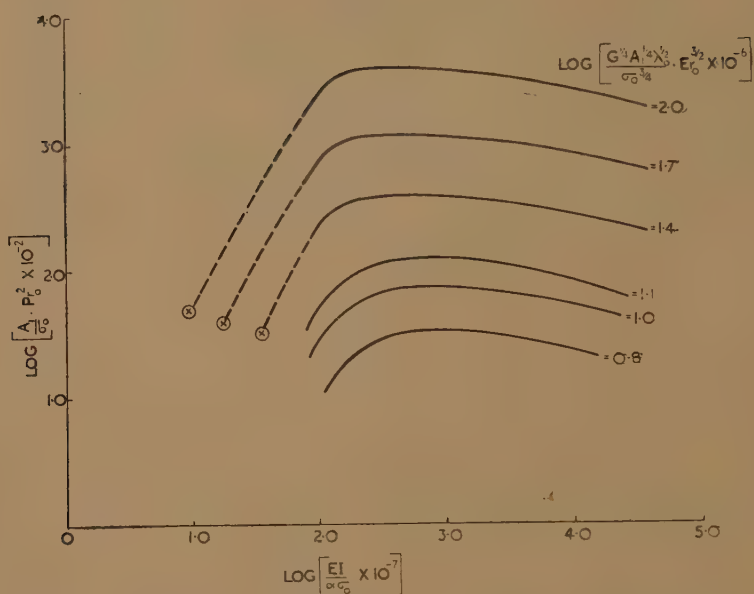
Arc Gradient $E \times 10^{-8}$	Loading $EI \times 10^{-7}$	$\log \left(\frac{EI}{\alpha \sigma_0} \times 10^{-7} \right)$	$\log \left(\frac{A_1}{\sigma_0} Pr_0^2 \times 10^{-2} \right)$	$\log \left(\frac{G^{1/4} A_1^{1/4} \lambda_0^{1/2}}{\sigma_0^{3/4}} \cdot Er_0^{6/4} \times 10^{-6} \right)$
<i>Lamp A.</i> Quartz tube. Arc length 6.0 cm. Internal diameter 25 mm. 4 mm. tungsten electrodes. P=0.76				
57	114	.87	1.69	2.11
63	63	.62	1.69	2.15
50	140	.96	1.69	2.05
<i>Lamp A.</i> P=0.665				
38	163	1.03	1.63	1.93
30	225	1.16	1.63	1.83
31	214	1.15	1.63	1.84
27	243	1.20	1.63	1.78
20	300	1.29	1.63	1.66
<i>Lamp B.</i> Glass tube. Arc length 16 cm. Internal diameter 30 mm. P=0.605				
88	41	0.43	1.84	2.5
81.5	53	0.54	1.84	2.46
75	60	0.59	1.84	2.43

These difficulties could to some extent have been removed by extending the numerical work to lower values of EI by making calculations with smaller values of $A_3 P^{3/2} / G^{1/2} E^2$. It was thought, however, that the verification obtained was sufficiently satisfactory without this further labour. The sets of experimental results were obtained which are given in Table I.

In these calculations, owing to the large thermal capacity of the electrodes employed; a total voltage drop of 40 V. at anode and cathode was assumed. To obtain the final columns, the helium constants estimated in the last section were, of course, used.

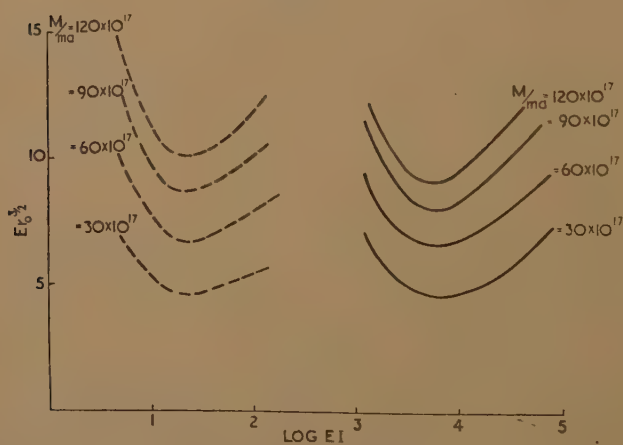
The helium curves in fig. 2 were extended—partly by extrapolation—as shown in fig. 4. The three crosses shown on the dotted extensions of the three upper curves were obtained from the smoothed results of Table I.

Fig. 4.



Full lines : calculated characteristics. Crosses : measured values.

Fig. 5.



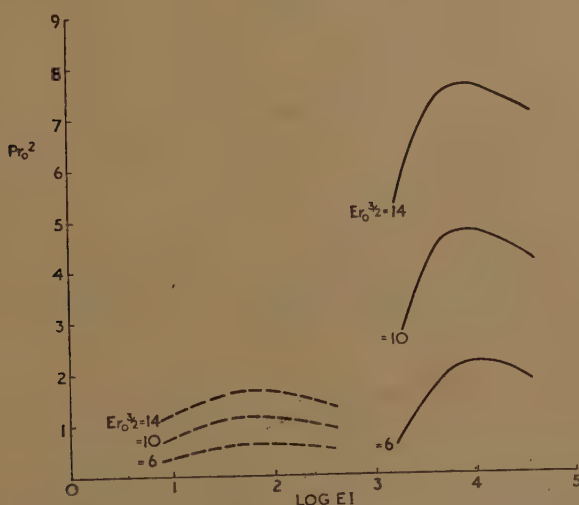
Arc voltage—arc loading (watts) characteristic at constant density ;
 ————— helium.
 - - - - - mercury vapour.

It appears that there is no disagreement between the results deduced from measured values using the estimated values of the constants and the calculated curves. The values for helium given in equation (4.5) were therefore assumed in the curves of the next section.

§ 6. CHARACTERISTICS FOR THE HIGH PRESSURE HELIUM DISCHARGE.

Using these values of the constants, the relations were derived between loading and arc gradient at constant density shown in fig. 5, that between pressure and loading at constant gradient in fig. 6, between loading and maximum arc temperature at constant gradient in fig. 7. In all these curves, the electrical units are volts and watts.

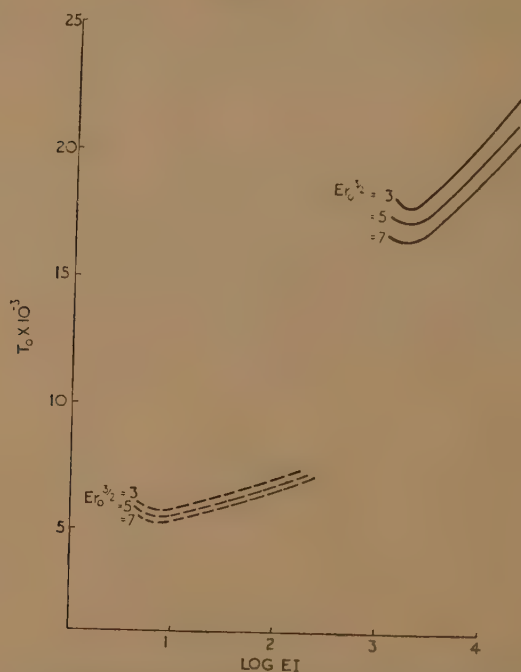
Fig. 6.



Pressure—arc loading (watts) characteristic at constant arc voltage ;
 ————— helium.
 - - - - - mercury vapour.

The outstanding conclusion is that the conditions corresponding to the state of operation normally obtained in the high pressure mercury discharge at the order of 10 to 100 watts per cm. of arc is not reached in helium until a loading of nearly 10,000 watts per cm. of arc is used. This is evident in all the figures, and calculations show also that at these loadings radiation efficiencies can be obtained comparable to those for mercury vapour. It is fairly certain in fact that, even with helium, a luminous efficiency approaching that of a black body at the optimum temperature, about 84 L/W, could in principle be reached; the suggestion from the present calculations, however, is that fantastically high loadings would be required. Fig. 7 is of interest in showing that at normal pressures, the operating temperature of the helium discharge is 15,000—20,000° C. instead of the 6000° C. of the mercury discharge.

Fig. 7.



Temperature—arc loading (watts) characteristic at constant arc voltage ;
 ————— helium.
 - - - - - mercury vapour.

REFERENCES.

- FRANCIS, V. J., 1946a, *Phil. Mag.*, **37**, 433; 1946b, *Ibid.*, 653; 1949, *Ibid.*, **40**, 435.
 ALDINGTON, J. N., *Trans. I.E.S.* (London). To be published.
 BRODE, R. B., 1930, *Phys. Rev.*, **35**, 504.
 KOLLATH, R., 1930, *Phys. Z.*, **31**, 985.

XCVIII. *Nuclear Transmutations Produced by Cosmic-Ray Particles of Great Energy.*—Part II. *Observations at High Altitudes by means of Free Balloons.*

By U. CAMERINI, T. COOR, J. H. DAVIES, P. H. FOWLER, W. O. LOCK,
H. MUIRHEAD and N. TOBIN.

The H. H. Wills Physical Laboratory, University of Bristol*.

[Received August 2, 1949.]

[Plates XVII.–XIX.]

SUMMARY.

In Part I. of the present series of papers (Brown *et al.* 1949), which we shall refer to as (I.), an account has been given of experiments with “electron-sensitive” photographic emulsions exposed to the cosmic radiation on the Jungfrauoch at 11,000 feet. In this second part we describe similar observations made at altitudes up to 130,000 feet by means of free balloons. In analysing the results, it will be convenient to adopt the same nomenclature for the classification of events of different types that was employed in (I.).

It is shown that, at $\sim 75,000$ feet, the radiation which leads to the production of “stars” is made up of charged and neutral particles of which the directions of motion are more widely dispersed than at lower depths in the atmosphere; and that high energy processes are relatively more frequent than at 11,000 feet. By making measurements of the deviations in the tracks of fast particles due to Coulomb scattering, it has been possible, in cases favourable for measurement, to give estimates of the energy of the particles, or to determine minimum values for this quantity. In four out of the seven cases which have been measured, the shower particles have been found to be less massive than the proton. It is shown that balloon flights give favourable conditions for studying the generation of π -particles in matter in the vicinity of the plates, and of estimating the rate of production of mesons in the energy range below ~ 60 MeV.

An event is described which appears to be due to the impact of a chlorine nucleus, ($z=17\pm 2$), with one of silver or bromine. The primary particle is almost completely dispersed into its component nucleons. Photo-micrographs are given of characteristic examples of the nuclear transmutations under discussion.

* Communicated by Prof. C. F. Powell, F.R.S.

EXPERIMENTAL.

DURING the exposures at high altitudes, the plates were enclosed in a Dewar vessel, of the type shown in fig. 1, in order to avoid submitting them to excessive changes of temperature. In some cases, the plates were arranged above or below lead, in the way indicated in the figure, in order to observe effects associated with the generation of secondary radiations in the lead.

Fig. 1.

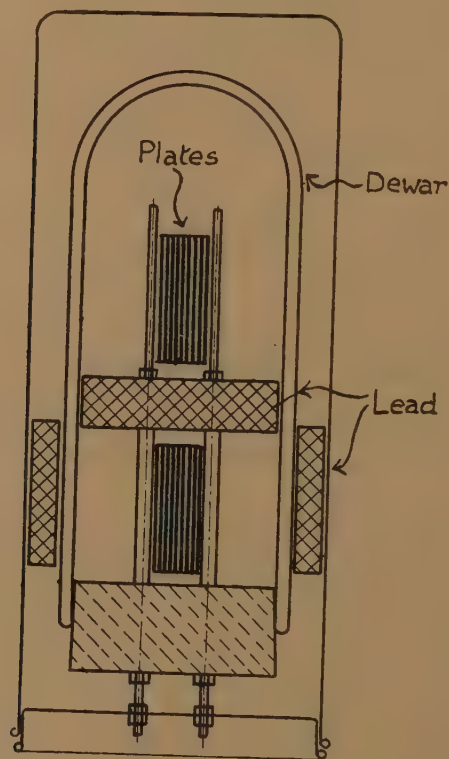


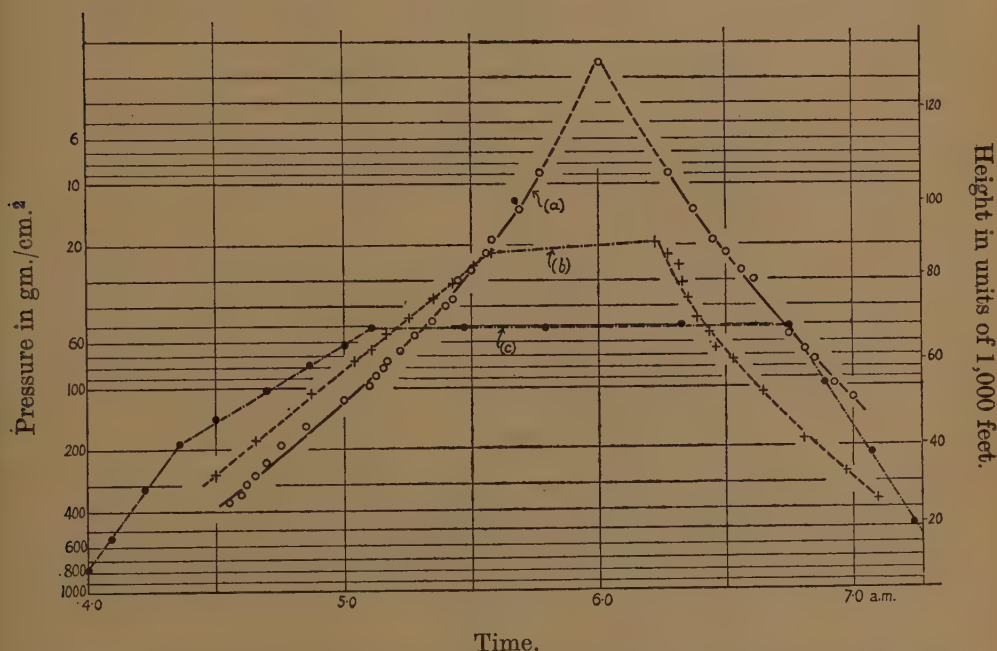
Plate assembly in Dewar flask, with lead blocks.

The gondola supported by the balloons carried "radio-sonde" equipment, so that the height of the assembly and its bearing from stations on the ground could be observed throughout the flight. Three typical curves showing the variation with time of the height of the balloons are reproduced in fig. 2. The flight corresponding to curve (a) in this figure was made with two 4 kg. balloons of natural latex, each inflated to give a net lift of 5 kg., the total load being 6 kg. It will be seen from the figure that the balloons rose at a mean velocity of 900 feet per min., until one of them burst. The remaining balloon then descended at a speed of 700 feet per minute.

In the flight corresponding to curve (b) in fig. 2, three 4 kg. balloons were employed, two of which were so inflated that, at ground level, they were just able to support the weight of the gondola. The third balloon was more completely inflated to provide lift, and provision was made to release it at 85,000 feet. This was done by completing an electrical circuit at this altitude by means of a "baro-switch". An electric current then flowed through a resistance wire which became hot and cut through the nylon cord attaching the balloon to the gondola.

This particular flight was made in conditions of good visibility, so that the ascent of the balloons could be observed visually. At 86,000 feet, the third balloon was seen to become detached from the gondola. The

Fig. 2.



Variation of altitude with time in three characteristic flights. The altitude is determined by means of a "baro-switch" of which the calibrations, before and after each ascent, are found to be consistent. The apparent increase in the rate of ascent at high altitudes in flight (a) is perhaps due to changes in the temperature of the instrument; but the large number of tracks of heavy nuclei observed in the plates exposed in this flight makes it reasonable to suppose that the plates were above 100,000 ft. for an interval of time approximately equal to that indicated in the figure.

remaining two balloons then floated at nearly constant altitude, until a second balloon burst some 40 minutes later, after which there was a rapid descent.

In the flight represented by curve (c), three 4 kg. balloons were employed which were so inflated that at ground level any two of them were just able to support the load. Each of the three balloons was provided

with a mechanical release, such that if a balloon burst, its debris would be jettisoned. This procedure has the advantage that the level flight is made at the maximum altitude reached by the least durable balloon; and it is this balloon which is jettisoned.

By making flights in conditions in which the high altitude winds are of low velocity and moving in favourable directions, and by ensuring that at least one balloon is still inflated when the gondola reaches the ground so that it acts a signal, we have succeeded in recovering our plates from more than 95 per cent of the ascents.

In order to estimate the intensity of the radiations producing stars of different types at a given altitude, we have adopted the following procedure: We assume that the intensity of the radiation which produces the stars with a given value of n_s varies with the depth of the atmosphere according to the relation $I=I_0e^{-x/k}$, where x is the mass, measured

TABLE I.

Star type	Value of K	
	Gross	Exponential
$\mathcal{N}_s \geq 0$	~ 170	~ 130
$\mathcal{N}_s \geq 3$	~ 135	~ 105
$\mathcal{N}_s \geq 6$	~ 120	~ 95

in gm. per cm.², of the overlying air. Alternatively, we employ the Gross transformation which takes account of effects due to the isotropic distribution of the directions of motion of the incoming particles at the top of the atmosphere:

$$I=I_0\left\{e^{-x/k}-\frac{x}{k}\int_{x/k}^{\infty}\frac{e^{-z}}{z}\cdot dz\right\}.$$

We can then calculate the numbers of days exposure at the Jungfraujoch to which any flight was equivalent, for any assumed value of k . The observed number of stars in the plates per cm.³ compared with the number produced in a given exposure at 11,000 ft., then allows us to determine k . In level flights of the type represented in fig. 2 (c), the contribution to the exposure made during the ascent and descent of the balloon is of the order of 20 to 30 per cent.

It is well known that there are serious objections to assuming an exponential law of absorption for the "star"-producing radiation, because the process of absorption is commonly not "catastrophic"; but it probably gives a sufficient close approximation of the corrections to be applied in our experimental conditions. In effect, the procedure allows

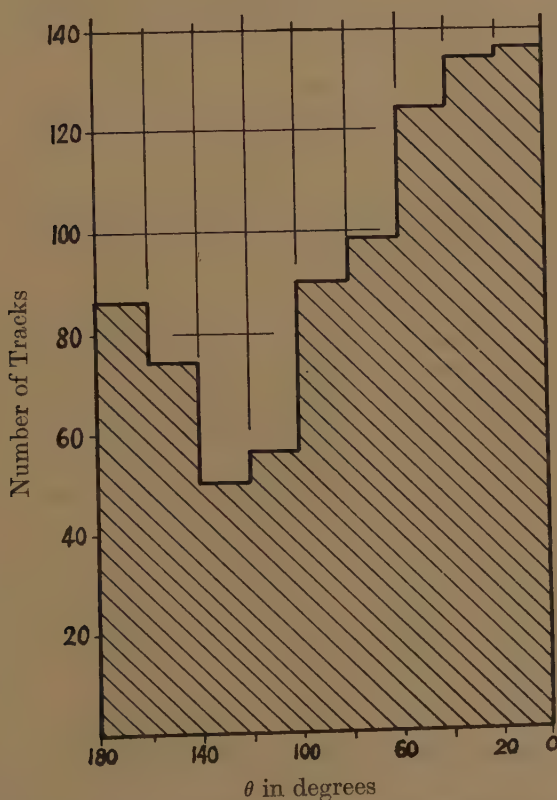
us to estimate the contribution to the time of level flight at maximum altitude to which the periods of ascent and descent are equivalent. Further, it allows us to compare our results with those of other workers who have analysed their observations in terms of an exponential law of absorption.

Table I. shows the absorption coefficients, deduced by the above methods, for the radiation producing stars of different types, from observations obtained in a number of balloon flights.

ORIENTATION OF THE TRACKS OF FAST CHARGED PARTICLES ASSOCIATED WITH "STARS".

We have determined the angular distribution of the directions of motion of all tracks of minimum grain-density, g_{\min} , associated with stars.

Fig. 3.

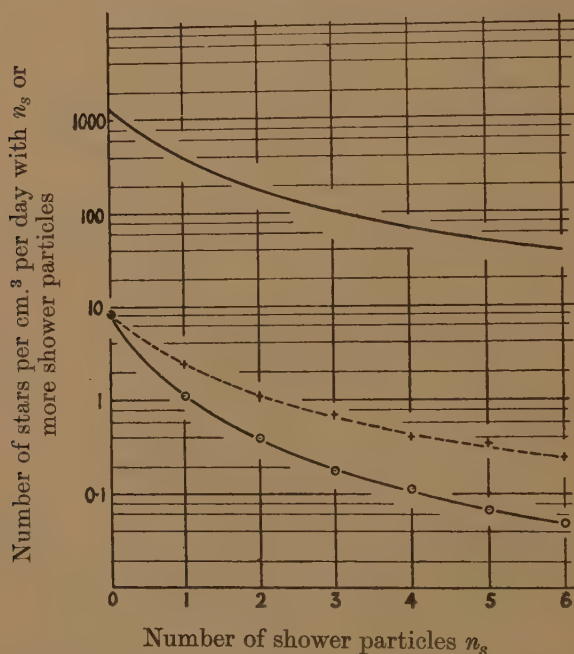


Distribution in the values of θ for all thin tracks associated with stars observed in a series of plates exposed at high altitude. θ is the angle between the vertical and the projection of the track on the surface of the emulsion.

produced in plates exposed at high altitudes, and the results are represented in fig. 3. This figure may be compared with the corresponding distribution observed at 11,000 feet—see (I.), fig. 4. It will be seen that at higher altitudes the degree of collimation of tracks near the vertical is less pronounced.

The same types of stars, both with and without "shower" particles, are observed at high altitudes as at 11,000 feet, but the proportion of "showers" and of "stars" with many branches is considerably greater at higher altitudes. Thus Table II. shows the numbers of stars of the different classes observed in the present series of observations. In fig. 4 we have plotted the rates of production of stars with different values of n_s (the number of shower particles) observed in the two series of exposures. The results clearly indicate the larger proportion of events produced by particles of high energy at the greater altitude.

Fig. 4.



Rates of production of stars with different values of n_s at 70,000 ft. and at 11,000 ft. For comparison, the dotted curve shows the results for 70,000 ft. normalized to correspond to a total rate of production equal to that observed at 11,000 ft.

A second important feature of the results at high altitudes is connected with the primaries of the events in which showers of fast particles are produced. At 11,000 feet, the "showers" are produced in approximately equal numbers by charged and uncharged primary particles, whereas at high altitudes the percentage of those events which are produced by charged particles is sensibly higher.

The relative frequency with which nuclear explosions are accompanied by different numbers of shower particles, n_s ,—each with a grain-density, g , less than $1.5g_{\min}$ —in plates exposed at high altitudes, is shown in fig. 5*a*. The form of the distribution is similar to the corresponding curve given

TABLE II.—Number of heavy particles N_h

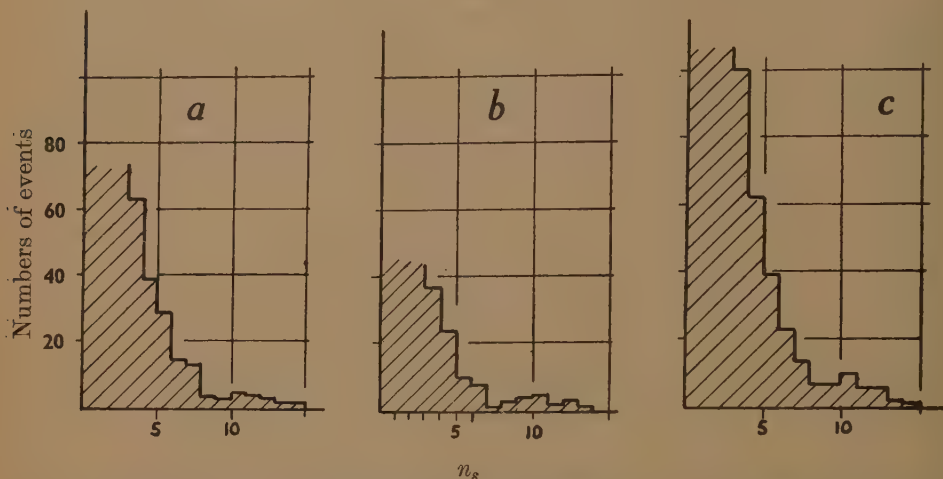
TABLE II.—Number of heavy particles N_h																										
Number of emitted "shower" particles																										
2	3	4	5	6	7	8	9	10	11	12	12	14	15	16	17	18	19	20	21	22	23	24	25	≥ 26	Total	
0^np	135	91	(75	47	26	21	16	9	5	9	2	1	5	2	1	0	1	1	1	1	1	..	447
1^np	5	7	14	13	8	11	9	3	2	2	0	3	0	0	2	1	0	0	1	1	1	82
2^np	5	9	6	4	6	5	8	4	2	6	0	1	1	1	..	3	..	1	1	1	59
3^np	5	6	11	9	6	4	4	5	0	2	5	0	1	2	1	1	0	1	1	..	1	65
4^np	1	2	0	1	1	1	0	0	1	4	2	..	3	1	0	..	1	9
5^np	..	2	1	1	..	1	..	2	..	1	2	1	2	..	1	22
6^np	1	1	1	2	1	3
7^np	1	1	1	3
8^np	1	1	1	1	1	1	1	1	..	6
9^np	1	2	1	1	1	3
10^np	1	1	..	1	1	1	1
11^np	1	1	1
12^np	1	0
13^np	1	1	2
	↓	0
	14 ⁿ	17 ^p	2

Number of emitted "shower" particles

To reduce the observed values to numbers per cm.³ per day at 70,000 ft., multiply by ~ 1.5 .
For the notation used see (I) page 877.

in (I.), fig. 12, which we reproduce for comparison in fig. 5*b*; and the results of the two sets of observations, taken together, are shown in fig. 5*c*.

Fig. 5.



Numbers of observed events with different values of n_s ; (a), results at high altitudes; (b), at 11,000 ft.; and, (c), both results taken together.

ENERGY OF THE SHOWER PARTICLES.

A "shower" particle is sometimes observed to be ejected in a direction nearly parallel to the surface of the emulsion, so that its track is of great length, in some cases longer than 10,000 μ .

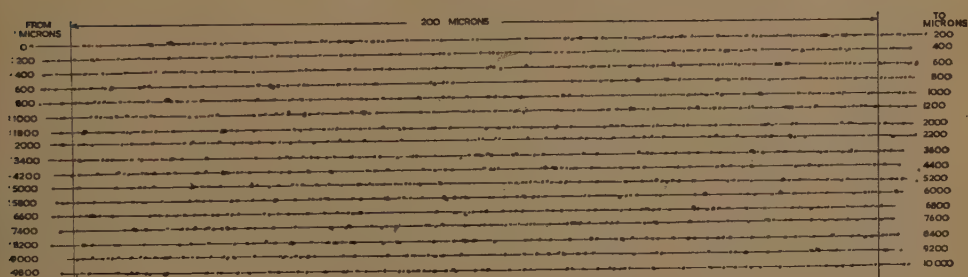
An estimate of the energy of such a particle can sometimes be obtained by observing the deviations in its track due to Coulomb scattering. We have made observations on ten examples of such long tracks, of which seven were produced by "shower" particles, and three were due to primary particles which were observed to produce a "shower". It is to be expected that the latter will be of such great energy that the Coulomb scattering in their tracks will not be detectable in the conditions of our experiments, any apparent deviation being due to instrumental and reading errors. Such observations allow us, however, to deduce lower limits for the energy of the corresponding particles and to make estimates of the errors for which allowance must be made in observations on the tracks produced by particles of lower energy such as those of a shower.

Fig. 6 shows a characteristic example of facsimile drawings, made with the aid of a projection microscope, of elements of the track of a fast primary particle, each of length 200 μ . In making the observations, we determine the mean direction of the track in successive elements of length 200 μ ; and the average value of the difference in direction of elements 800 μ apart. Such observations give us a measure of the quantity $m_0\beta^2c^2/\sqrt{1-\beta^2}$, which is equal to pv , where p is the momentum of the particle, and v its velocity (Goldschmidt-Clermont *et al.* 1948; Davies *et al.* 1949.)

Table III. shows the mean values of the deviations determined by this method; and the corresponding values of the energies of the particles for different assumed values of their rest-mass. It is a satisfactory feature of the observations that in three cases we have found the "primary" (P) and one of the "shower" (S) particles of the same event to be suitable for measurement. The very small deviations observed in the case of the "primary" particles gives us confidence that the greater values found for the shower particles are real, and that the corresponding values for energy are not seriously in error.

The tracks of which details are given in Table III., are distinguished from others only by their great length, which is a consequence of their favourable directions of ejection. It may be objected that tracks of lower energy would be more scattered and would therefore be less likely

Fig. 6. •



Facsimile drawing of elements, 200μ long, of the track of a shower particle of total length $10,000\mu$. The track is rectilinear to within about 0.6° , and the energy of the particle which produced it must have been greater than 1.5×10^9 e.v.

to remain in the emulsion, so that there would be a tendency to select for measurement only the tracks of particles of the greatest energy. If, however, the shower particles are mesons or heavier particles, they must necessarily have an energy of at least 5×10^7 e.v. if they produce tracks with a grain-density near the minimum value. The angles of scattering to be expected for such particles are so small that the deviations will produce no appreciable effect on the proportion of them which escape from the emulsion. This conclusion would not apply to the tracks of electrons with energies less than 40 MeV; but there are reasons for believing that electrons of such low energy constitute only a small proportion of the shower particles.

The above considerations make it reasonable to assume that the tracks on which we have made measurements constitute a random sample of those shower particles which are ejected in directions making small angles with the vertical. This limitation follows from the fact that more obliquely directed particles will, for purely geometrical reasons, have only a small probability of producing a track of great length. Now, at 11,000 ft.,

TABLE III.

Plate No.	Type of Event	Length of Track (microns)	Mean deviation in degrees ($t=800 \mu$)	Observed grain-density g/g_{\min}	Calculated values of energy and grain-density for various assumed values of rest-mass.						Angle to "primary" degrees	Nature of Particle	
					$m=m_e$			$m=300 m_e$					$m=m_p$
					Energy (MeV.)	g/g_{\min}	Energy (MeV.)	g/g_{\min}	Energy (MeV.)	g/g_{\min}			
					Energy (MeV.)	g/g_{\min}	Energy (MeV.)	g/g_{\min}	Energy (MeV.)	g/g_{\min}			
KE 28	$8_p(\mathcal{N}_h=15)$	P	0.05	0.98	2000	1.0	1850	1.0	1500	1.0	uncertain		
		S	0.06	0.98	1500	"	1350	"	1100	"	uncertain		
		S	0.30	1.09	310	"	220	"	180	2.0	$m < m_p$		
KF 2	$5_p(\mathcal{N}_h=13)$	P	0.066	1.06	1400	"	1250	"	900	1.0	uncertain		
		S	0.15	1.06	600	"	500	"	360	1.4	$m < m_p$		
KF 19	$6_p(\mathcal{N}_h=8)$	P	0.032	0.98	2900	"	2750	"	2300	1.0	uncertain		
		S	0.11	0.86	850	"	740	"	550	1.1	uncertain		
KE 9	$10_p(\mathcal{N}_h=9)$	S	0.22	1.0	420	"	280	"	240	1.7	$m < m_p$		
KE 18	$3_n(\mathcal{N}_h=6)$	S	0.22	1.12	420	"	280	"	240	1.7	$m < m_p$		
KF 15	$5_p(\mathcal{N}_h=9)$	S	0.12	1.0	760	"	610	"	450	1.2	uncertain		
" Grey " Tracks													
KE 27	$2_n(\mathcal{N}_h=8)$	S	0.36	1.7	260	1.0	180	1.0	150	2.1	probably proton		
KF 4	$6_p(\mathcal{N}_h=20)$	S	0.13	1.6	230	"	160	1.06	130	2.1	probably proton		
KF 24	$5_p(\mathcal{N}_h=7)$	S	0.126	1.7	240	"	170	1.13	130	2.1	probably proton		
KE 10	$4_n(\mathcal{N}_h=7)$	S	0.6	1.75	50	"	30	2.0	20	>5	$m < m_p$		

Comparison of observed values of the grain-density in the tracks of "relativistic" particles associated with stars, with calculated values for different assumed values of the rest-mass.

the tracks of charged primary particles usually only make small angles to the vertical. It follows that the long tracks of shower particles will commonly be inclined at only a small angle to that of the primary particle which produced them. Furthermore, it is reasonable to suppose that the average energy of such shower particles is greater than that of similar particles of which the direction of motion are more widely dispersed.

A METHOD OF DETERMINING THE REST-MASS OF THE SHOWER PARTICLES.

The results given in Table III. show that some of the shower particles have energies $\sim 3 \times 10^8$ e.v. This value is small enough to allow us to distinguish between the possibilities that the particles in question are mesons or protons. The track of a proton has a grain-density, g , significantly greater than g_{\min} only if its energy is less than 5×10^8 e.v., whereas above 10^8 e.v., a π -meson for example, will produce a track with $g = g_{\min}$. It follows that for particles of energy less than 5×10^8 e.v., the observation both of the scattering of the particle and of the grain-density in its track will allow us to show whether or not it is less massive than a proton and, in favourable cases, to determine its rest-mass.

The above method is analogous to that commonly employed in experiments with expansion chambers operated in magnetic fields. In such experiments, the velocity of a particle is deduced from its specific ionization as determined by drop-counts, and its momentum from the curvature of its trajectory, whereas in the photographic method, the grain-density gives a measure of the velocity and the observed scattering the quantity pv .

In the case of four of the tracks of "shower" particles of which details are given in Table III., we are able to conclude that the particles in question were less massive than the proton. In the three other cases, it is not possible to draw any conclusions.

The method is particularly well adapted to a study of the nature of the particles producing grey tracks and we have applied it in three cases suitable for measurement. We have thus been able to show that the corresponding particles were protons.

PRODUCTION OF π -PARTICLES AT GREAT ALTITUDES.

It is well known that many of the slow π -particles observed at mountain altitudes are locally generated, either in the material of the photographic plates themselves or in the other matter in their immediate vicinity. In the case of the balloon flights, the concentration of matter round the plates is greatly reduced, and we should therefore expect to observe a smaller number of π -particles relative to the number of stars. We have examined the plates exposed at the two altitudes in order to prove the correctness of this view, and the results are shown in Table IV.

The results given in Table IV. show the number of events actually recorded by the observers and uncorrected for geometrical and other effects. The numbers of mesons actually stopping in the emulsion must be greater than those observed, and the last line in the Table gives estimated values for the true ratios.

TABLE IV.

Observed number	11,000 ft.	$\sim 80,000$ ft.	
		with lead	without lead
$\pi + \sigma$ —mesons.....	566	48	72
stars ($\mathcal{N}_h \geq 3$)	5904	892	2029
$\mathcal{N}_\pi/\mathcal{N}_{st}$ —observed.....	0.096	0.054	0.035
$\mathcal{N}_\pi/\mathcal{N}_{st}$ —estimated true ratio....	0.11	0.062	0.040

The results obtained at high altitudes have been divided into two groups according to whether or not lead blocks were disposed round the plates in the way shown in fig. 1.

The results summarized in Table IV. show that the ratio, of the number of π^\pm -particles to the number of stars is about three times smaller, for the plates exposed without lead in the balloon flights, than in the plates exposed at 11,000 feet. The rapid increase in the number of mesons produced by even small thicknesses of lead is also well displayed by the results.

Using the numerical results given in Table IV., we have made an estimate of the rate of production, in lead at $\sim 80,000$ feet, of mesons with an energy less than ~ 80 MeV. We assume that the difference between the observed numbers of π -particles, in the experiments with and without lead, are due to the generation of mesons in the lead. These mesons will be brought to rest in the plates if they are emitted in suitable directions, and if their energy is less than about 60 MeV.

The calculations lead to the conclusion that the number of mesons produced per star, in the specified range of energies, is 0.06. This value may be compared with the rate of production of shower particles in the emulsion, *viz.* 0.05 per star.

Our results therefore show that a large fraction of the mesons created by the interaction of cosmic radiation with solid materials, are emitted with energies less than 60 MeV. It must be emphasized that the precise value of the maximum energy of the mesons recorded in these experiments is subject to some uncertainty. Further experiments are in progress and a more detailed analysis will be given in a later communication.

SECONDARY NUCLEAR INTERACTIONS PRODUCED BY SHOWER PARTICLES.

It has sometimes been observed, in the present experiments, that a charged particle emerging from a nuclear explosion with a velocity in the relativistic region, can interact with a nucleus and produce a second disintegration. Photo-micrographs of two events of this type are shown

in Pls. XVII. and XVIII. In Pl. XVII., the first star, A, is of type 2_n , and in Pl. XVIII., of type 1_n . A similar event has been described by Herz in which the first "star" is of type 1_n .

Observations of this character have a bearing on the problem of the identity of the shower particles. By studying the average length of their path in the emulsion before they make a nuclear collision, we may obtain some insight into the strength of the interaction of the particles with nucleons. The approach is similar to that in experiments with expansion chambers in which studies are made of the penetrations of lead plates by the particles of a shower. Our present observations are, for the present, not sufficiently extensive to allow us to draw any conclusions, but it is perhaps interesting to note that the total observed length in the emulsion of the "thin" tracks emerging from stars of type 1_n and 2_n , 1_p and 2_p is 60 cm, corresponding to the passage of the particles through an absorber of mass ~ 240 gm. per cm.² The particles giving rise to these tracks have been observed to produce three nuclear collisions. On the other hand, the total length of the track of fast particles emitted from stars with higher values of n_s , none of which have been observed to produce a disintegration, is 65 cm. This point will be discussed in more detail in Part III.

SHOWERS OF FAST PROTONS PRODUCED BY HEAVY PRIMARY COSMIC-RAY PARTICLES OF GREAT ENERGY.

The plates exposed at high altitudes contain many examples of the tracks of heavy nuclei of the cosmic radiation (Freier *et al.* 1948), and Pl. XIX. shows an example in which one of these particles has interacted with a nucleus of silver or bromine. The event appears to provide a striking example of a nuclear collision of the type described by Bradt and Peters (1949), which they distinguish as class (c).

The track of the particle (1) in the figure clearly displays the large number of δ -rays characteristic of heavy nuclear fragments moving with great velocity. The track can be followed, without a change of direction greater than 1° , through the emulsion of four plates in succession, and the material which the particle is thus observed to have traversed has a mass of about 8 gm. per cm.². The number of δ -rays per unit length of the track is nearly constant over the whole of the observed path, and it follows that the particle was moving with a velocity closely approaching that of light. By determining the number of δ -rays per unit length of the track, the charge on the particle can be determined and the result thus obtained is $17 \pm 2e$, where e is the electronic charge.

Several examples have now been found, in this and other laboratories, of the ejection—during the explosive disintegration of silver and bromine nuclei—of heavy nuclear fragments with an energy of the order of 200 MeV. In the present instance, however, the energy of the heavy particle was greater than 10^4 MeV. Further, if the particle approached the "star", it was moving downwards at an angle of 60° to the vertical,

PLATE XVII.

A particle producing the track p_1 interacts with a nucleus at A and produces a nuclear explosion, of type $2p$ ($\mathcal{N}_h=8$), from which two "relativistic" particles of charge e emerge. One of the fast secondary particles, p_2 , emerging from the disintegration, A, makes a second nuclear explosion at B, of type $O_p(\mathcal{N}_h=9)$. In Part III. of the present series of papers, it will be shown that of the events of type $2p$, about 75 per cent correspond to the passage of a fast proton through a nucleus and the creation of a single charged "shower" particle; and that of the events of type $O_p(\mathcal{N}_h \geq 9)$, the majority correspond to the passage through a nucleus of an incident proton, which, through charge exchange, emerges as a neutron. The most probable explanation of the event therefore is that it corresponds to the passage of a fast nucleon through two nuclei in succession. In this case, the photograph gives a pictorial representation of one of the main modes of loss of energy, in passing through the atmosphere, of the primary protons with energy in the interval from ~ 3 to $\sim 10 \times 10^9$ e.v.

PLATE XVIII.

An event of a similar character to that shown in Pl. XVII. We attribute it to an incident neutron, which produced the star of type $I_n(\mathcal{N}_h=7)$ at A, and emerged as a proton to produce the star at B of type $O_p(\mathcal{N}_h=2)$. The interpretation depends on the analysis, which will be given in Part III., which indicates that 80 per cent of stars of type I_n are due to the passage of a proton through a nucleus with exchange of charge.

PLATE XIX.

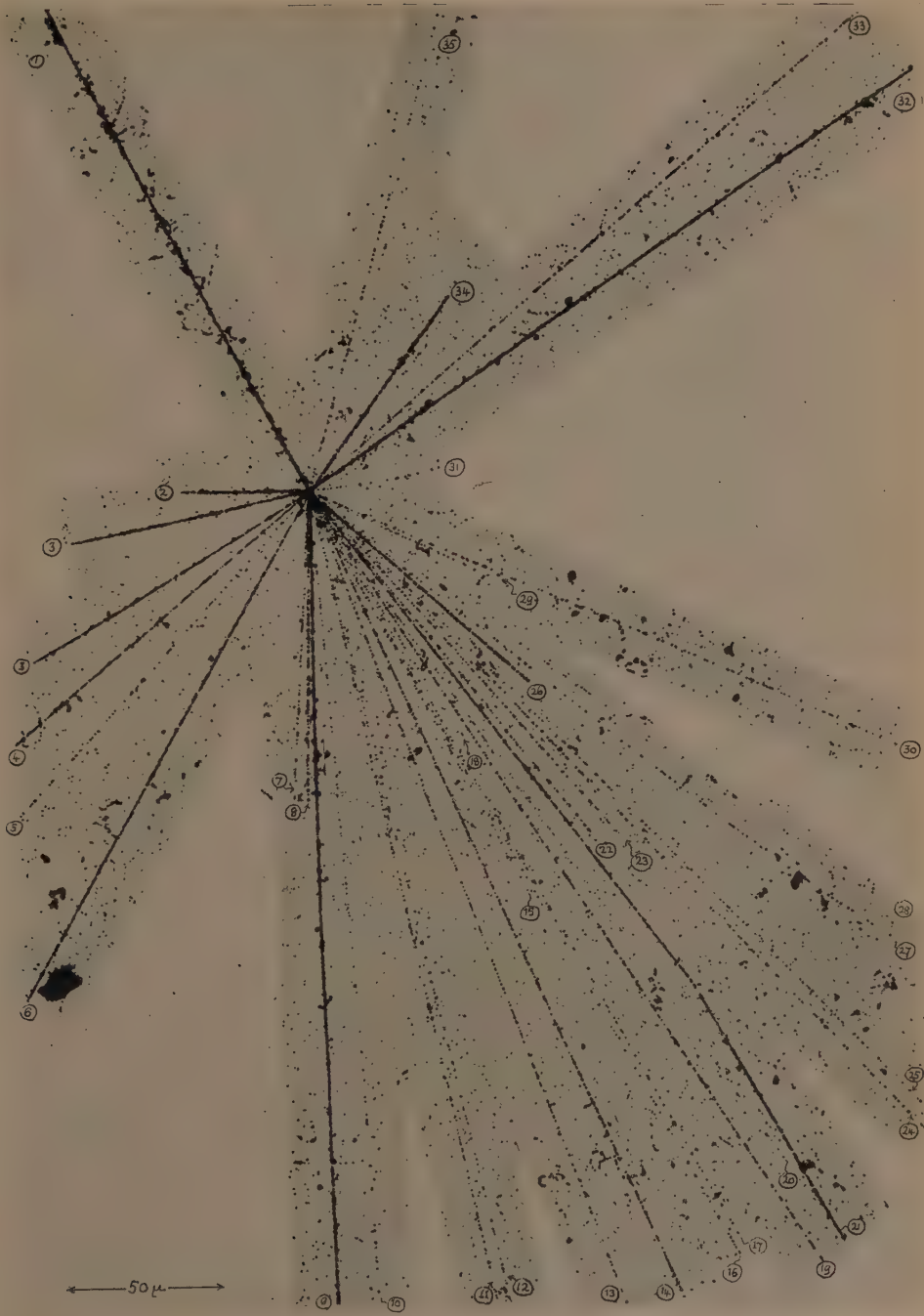
The track (1) can be followed through the emulsion of four plates, and the number of δ -rays per unit length, which is nearly constant, indicates that the charge of the particle is $17 \pm 2e$. The event is therefore interpreted as due to the impact of a chlorine nucleus, with one of silver or bromine in the emulsion. As a result of the impact, the nucleons of the incident nucleus appear to have been almost completely dispersed, so that they emerge as protons and neutrons moving in directions embraced by a narrow cone. Most of the more heavily ionizing particles are due to the evaporation of the struck nucleus. Track (32) is that of a Li nucleus; (9), that of an α particle; and (4), that of a fast particle which produces a second disintegration at a point outside the field of view represented in the photograph.



Observer : M. STOTT.



Observer: MARY JONES.

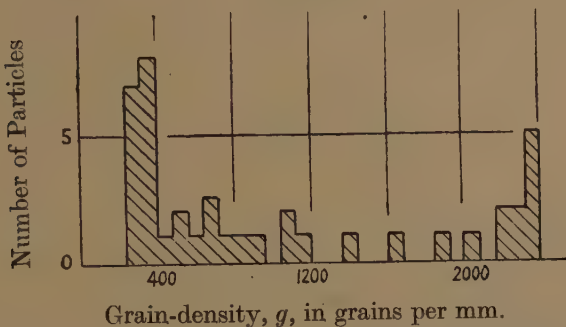


Observer : B. HULBERT.



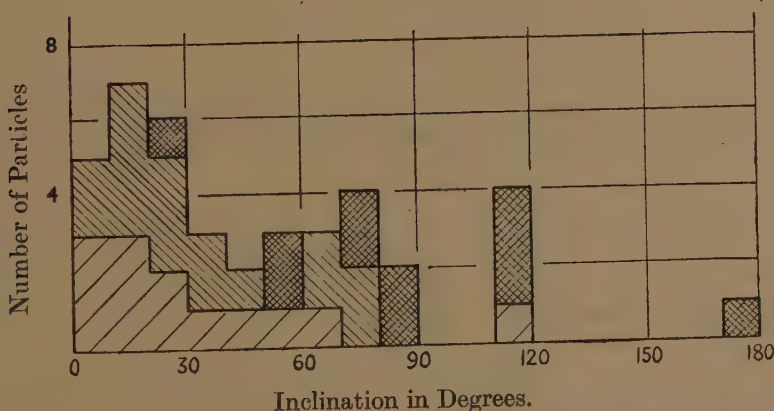
and its line of motion was very close to the "axis" of the cone of fast particles emerging from the disintegration. It is certain, therefore, that the particle formed one of the heavy nuclei of the "primary" cosmic radiation; that it collided with the nucleus to produce the observed disintegration, and did not appear as one of its products.

Fig. 7.



Histogram showing the values of the grain-density in the tracks of all particles associated with the event shown in Pl. XVIII.

Fig. 8.



Distribution in the values of the inclinations of the directions of motion of all particles associated with the event shown in Pl. XVIII., to that of the primary particle. Values for tracks with minimum ionization shown thus :

☐ ; "grey" tracks, thus : ▨ ; and "black" tracks, thus : ▩

A histogram showing the values of the grain-density, g , in all the tracks associated with the event, is given in fig. 7. It will be seen that fourteen of the secondary particle have a specific ionization less than 1.3 times the minimum value for particles of charge $|e|$. In fig. 8, we show the distribution in the direction of motion of the secondary particles relative to that of the heavy fragments assuming the latter to have approached

the disintegrating nucleus. Most of the tracks of the "fast" particles were ejected in directions making angles less than 30° with that of the assumed "primary" and the arithmetic sum of the charges carried by these particles was $12e$. These observations suggest that the event corresponds to the interaction of a fast chlorine nucleus with one of silver or bromine; and that as a result of the collision the nucleons of the incident particle have been almost completely dispersed so that they appear in the form of a narrow jet. Further evidence in support of this view will be given in Part III.

REFERENCES.

- BROWN, CAMERINI, FOWLER, HEITLER, KING and POWELL, 1949, *Phil. Mag.*, **40**, 862.
 GOLDSCHMIDT-CLERMONT, KING, MUIRHEAD and RITSON, 1948, *Proc. Phys. Soc.*, **61**, 183.
 DAVIES, LOCK and MUIRHEAD, *Phil. Mag.* (in course of publication).
 FREIER, NEY, OPPENHEIMER, BRADT and PETERS, 1948, *Phys. Rev.*, **74**, 213.
 BRADT and PETERS, 1948, *Phys. Rev.*, **74**, 828.
 HERZ, private communication and 1949, *Phys. Rev.*, **75**, 1779.

XCIX. *Experiments with Nuclear-Track Emulsions*
Sensitive at Minimum Ionizing Power.

By A. C. COATES, A.R.C.S., and R. H. HERZ, D.Phil.,
 Research Laboratories, Kodak Ltd., Harrow, Middlesex*†.

[Received August 10, 1949.]

[Plates XX.-XXIII.]

SUMMARY.

The angular distribution of straight-line tracks due to high energy cosmic ray particles at sea level and their number per cm^2 per day at sea level and below an absorber (equivalent to 60 m. of water) have been investigated. Slow electron tracks are recorded at a rate of 10^4 per cm^2 per day in a 200 micron thick emulsion. The sensitivity of the emulsion to particles at minimum ionization does not change at low temperature ($-80^\circ\text{C}.$), although a decrease of 24 per cent was found for X-ray exposures at $-80^\circ\text{C}.$ as compared with room temperature.

A number of pair formation tracks produced by a 25 MeV. synchrotron have been observed. In all cases but two, one particle (presumably the

* Communicated by E. W. H. Selwyn.

† Communication No. 1281 H from the Kodak Research Laboratories.

positron) stops in the emulsion before slowing down. Two examples of elastic collisions between particles at minimum ionization and electrons and the formation of a multiple star are discussed.

INTRODUCTION.

SINCE the first announcement of nuclear track emulsions (Berriman 1948) (Kodak NT4), capable of recording tracks of charged particles of minimum ionizing power, investigations have been made in order to establish some of the characteristic features of such emulsions. This note describes a number of observations made during tests on the photographic behaviour of these plates.

HIGH-ENERGY PARTICLES.

High-energy cosmic-ray particles of minimum ionizing power are recorded in NT4 plates as straight-line tracks, up to several thousand μ long, of small grain density. These tracks are almost straight, because the Coloumb-scattering of the particles is small. Using the processing conditions recommended for these plates, the mean grain density along the tracks is 45 grains per 100 μ length, that is, about 20 per cent of the grains are affected if particles of minimum ionizing power pass through the emulsion.

From the known rate of energy loss of particles in matter it follows that if they are protons their energy must be at least 1800 MeV. for them to be at minimum ionizing power. Similarly, if they are mesons they must have an energy of 200–300 MeV., depending on the mass of the meson concerned. However, an electron producing minimum ionization and having an energy of the order of 1 MeV. would not be recorded as a straight line, because of the increased small angle scattering it suffers. An electron track showing a mean angle of scatter of $1/2^\circ$, measured on a cell length of 100 μ corresponds to an energy of approximately 80 MeV.; a scatter of 1° corresponds to 40 MeV.; and of 2° to 20 MeV.

The zenith-angle distribution of straight-line tracks has been determined by placing plates 200 μ thick in the vertical position on the roof of the laboratory. Only those tracks lying in the plane of the emulsion have been counted, and fig. 1 shows a histogram of 350 counts, which agrees well with the $\cos^2 \theta$ distribution as previously shown by counter measurements (Clay *et al.* 1938, Kolhörster *et al.* 1932).

The number of high-energy particles crossing a horizontal area of 1 cm.² per day has been measured at sea level and in the underground laboratory (ceiling equivalent to an absorber of 60 m. of water) of Birkbeck College, University of London. The maximum angle of dip which allows easy recognition of straight-line tracks at minimum ionization was found to be 30° . Tracks forming larger angles of dip have not been counted, but their number has been computed from the angular distribution above,

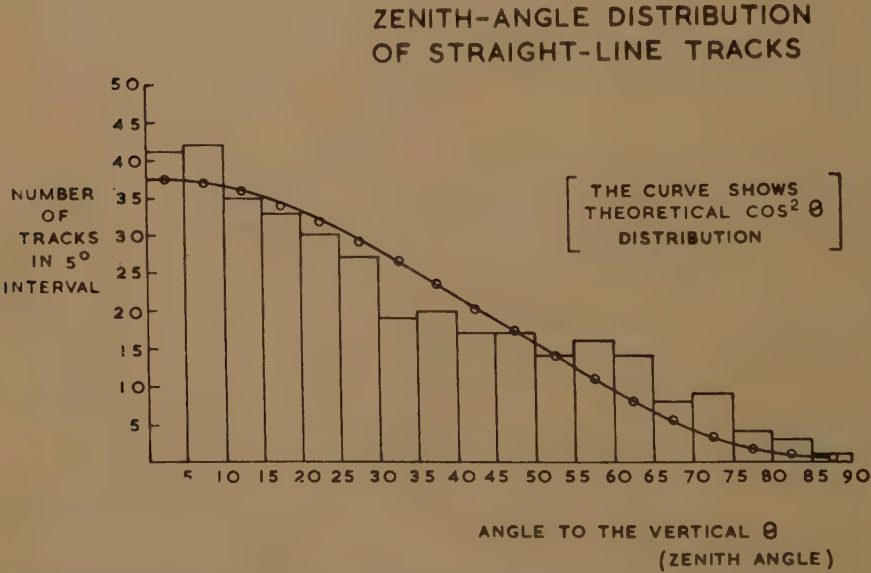
The minimum energy of particles counted is approximately 20 MeV. This figure has been obtained from measurements of small angle scatter. (Goldschmidt-Clermont *et al.* 1948.)

In the Table below the results are quoted as the number of straight-line tracks obtained per horizontal area of 1 cm.² per day, at sea level and in the underground laboratory. It is estimated that about 80 per cent of the tracks recorded at sea level are due to high-energy mesons.

Storage Conditions.	No. of Tracks per cm. ² /day.
Sea level	840
Underground laboratory.....	56

The figure obtained at sea level is 20 per cent less than that published by Greisen (1942) using counter measurements, but his results include slow mesons which do not produce straight-line tracks at minimum ionization.

Fig. 1.



Zenith angular distribution of straight-line tracks in the plane of the emulsion.

Only 7 per cent of the number of tracks at sea level has been observed in the underground laboratory, the ceiling of which absorbs all particles of energies below 10,000 MeV. Dr. E. P. George, Birkbeck College, using counter measurements, has found the same percentage reduction (private communication). All the tests were made over three exposure periods, 4, 10 and 21 days ; and since the number of tracks counted was found to be proportional to the exposure time, the latent image fading appears to be negligible within this time,

LOW-ENERGY ELECTRONS.

In addition to the tracks due to the high-energy particles, a great number of wavy tracks due to slow electrons (energies up to a few MeV.) are continuously recorded in the emulsion. These are caused partly by cosmic rays and partly by electrons from the decay of radioactive traces in the surrounding air. The number of short electron tracks recorded in a 200μ thick emulsion has been found to be of the order of 10^4 per cm^2 per day. Some preliminary experiments have been carried out by suspending the plates in a glass vessel at reduced pressure (1 mm. Hg), in order to reduce the influence of the radioactivity of air. A reduction of only 30 per cent of the number of slow electron tracks has been observed at this pressure. Further experiments are being made in order to determine the relative contribution of the radioactivity of the air and of cosmic rays.

INFLUENCE OF TEMPERATURE ON SENSITIVITY.

It was first believed that the sensitivity of the emulsion to high-energy particles could be reduced temporarily by storing NT4 plates at low temperature. No significant difference in grain density of minimum ionization tracks has been observed between tracks recorded at room temperature and at -80°C . (solid CO_2). The sensitivity of NT4 plates exposed to 130 kV. X rays, however, has been found to decrease linearly over a temperature range of $+18^\circ\text{C}$. to -80°C . At the latter temperature the sensitivity is 76 per cent of that at room temperature. This effect is partly attributable to the high ionizing low-energy electrons released by X rays. Similar results with high-ionizing particles have been obtained by Castle and Webb, of the Eastman Kodak Research Laboratories (private communication), and Cosyns *et al.* (1949).

PAIR FORMATION.

An electron-positron pair caused by cosmic rays has been observed in these laboratories on a photographic plate for the first time (see fig. 2, Pl. XX.). More recently several examples of pair formation have been recorded in plates exposed to 25 MeV. X rays, using the synchrotron at Telecommunication Research Establishment, Great Malvern. Similar observations have also been made by Mr. D. T. King, University of Bristol, using the same plates exposed to the radiation of the synchrotron (private communication). Fig. 3*a* (Pl. XX.) and fig. 3*b* (Pl. XXI.) show various examples of pair-formation tracks created by 25 MeV. X rays. In all cases but two, one of the particles stops in the emulsion before slowing down; this particle is thought to be a positron which appears to be annihilated during its flight. Although it appears in most cases that the energy is equally shared between the two particles, fig. 3*a* (Pl. XX.) shows one example in which one particle (presumably the positron) has considerably higher energy than the other, as seen from the different amount of Coulomb-scattering.

An attempt has been made to evaluate the energies of three events by measurements of the Coulomb scattering. The results of two cases, No. III. and VI. in fig. 3*b* (Pl. XXI.), are tabulated below :—

<i>Event III.</i>	<i>Event VI.</i>
14.5 MeV. top track.	6.9 MeV. top track.
6.9 MeV. bottom track.	8.6 MeV. bottom track.
<hr/> Total energy 21.4 MeV.	<hr/> Total energy 15.5 MeV.

These values do not include the rest mass of the particles. The energies evaluated of the pair event shown in the photomicrograph, fig. 3*a* (Pl. XX.), lead to an energy value for the straight-line track which is about four times higher than the expected figure.

The tracks of all pairs produced by the synchrotron exposure run very nearly parallel and in the direction of the primary X-ray beam. No significant difference has been found in the grain density between positron and electron tracks. The angle separating the two tracks was found in all cases of the order of $1-2^\circ$.

ELASTIC COLLISIONS.

Two mosaics of photomicrographs shown in figs. 4 and 5 (Pl. XXII.) illustrate collisions of high-energy particles of minimum ionization with electrons. The angular deviation from the straight-line track at the point of collision can be readily seen in fig. 4 (Pl. XXII.). An attempt has been made to determine the energy of the fast particle from relativistic momenta relations, fig. 5 (Pl. XXII.). The energy of the knock-on electron has been derived from the energy-range relationship for electrons in the emulsion, and is about 175 kV. Taking the deviation from the straight line at the point of collision as 0.5° and the angle between the two tracks as 55° , it follows that the energy of the fast particle is 250 MeV. if it is an electron, or 50 MeV. if it is a π meson, and 60 MeV. if it is a μ meson. For a proton-electron collision calculation shows the energy of the incident proton to be 12 MeV. A proton of this energy would not have reached minimum ionizing power and can therefore be excluded.

DISINTEGRATION BY SECONDARY PARTICLE AT MINIMUM IONIZATION.

Several stars have been observed which seemed to have been initiated by particles of minimum ionization and the event shown in fig. 6 (Pl. XXIII.) gives evidence that such particles are capable of initiating stars and also that they may be of secondary origin. This event has been obtained in an NT4 plate which was exposed to cosmic rays on the Jungfrauoch. Star A, which has five heavily ionizing particles, has been initiated presumably by a neutral particle. One of the particles at minimum ionization has a range of 1600μ in the emulsion, with no change in grain density along its track, and produces a second star, B, of 12 particles.

Fig. 2.

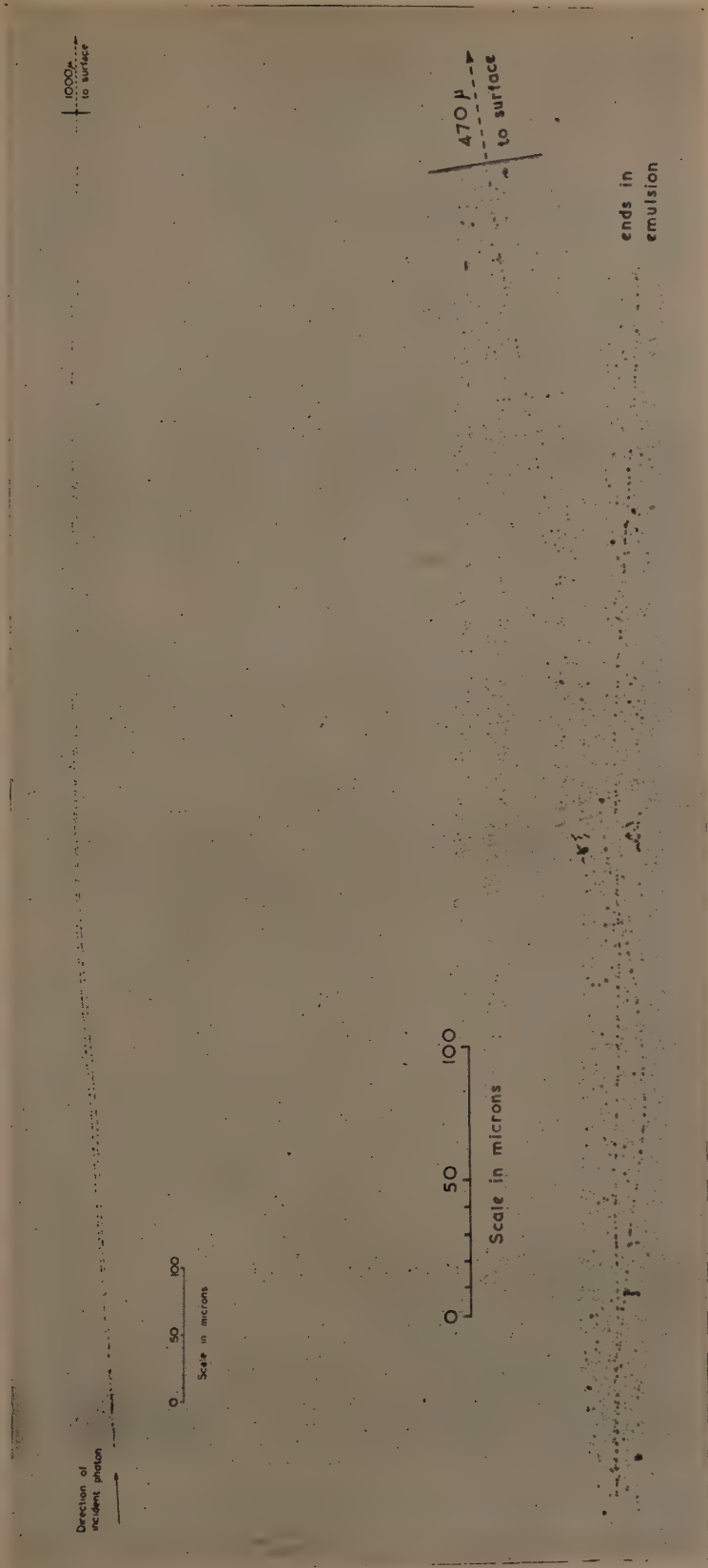


Fig. 3(a).

Fig. 2. Pair-formation tracks due to cosmic rays. Both particles have an energy of 18 MeV. as found from Coulomb scatter. It is believed that the track which ends suddenly in the emulsion is due to a positron. (Drawing made with projection microscope.) (Observer: Miss E. Herxheimer, of the Kodak Research Laboratories.)

Fig. 3(a). Pair formation tracks produced by exposure to X-rays from a 25-MeV. synchrotron. (Photomicrographic mosaic.)

FIG. 3 (b).

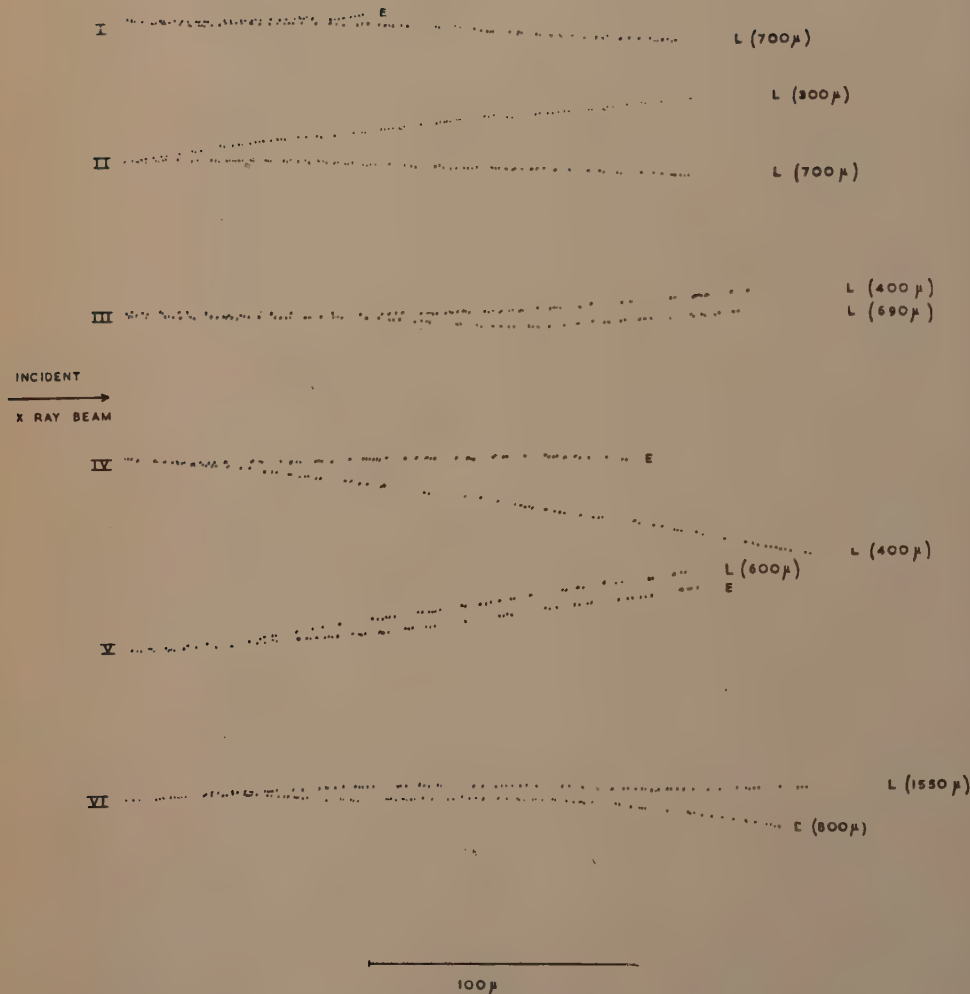


Fig. 3 (b). Tracks of several electron-positron pairs. (Drawings made with projection microscope.)

E ends in emulsion.

L leaves emulsion.

The figures in brackets denote the lengths of the tracks observed in emulsion.

FIG. 4.

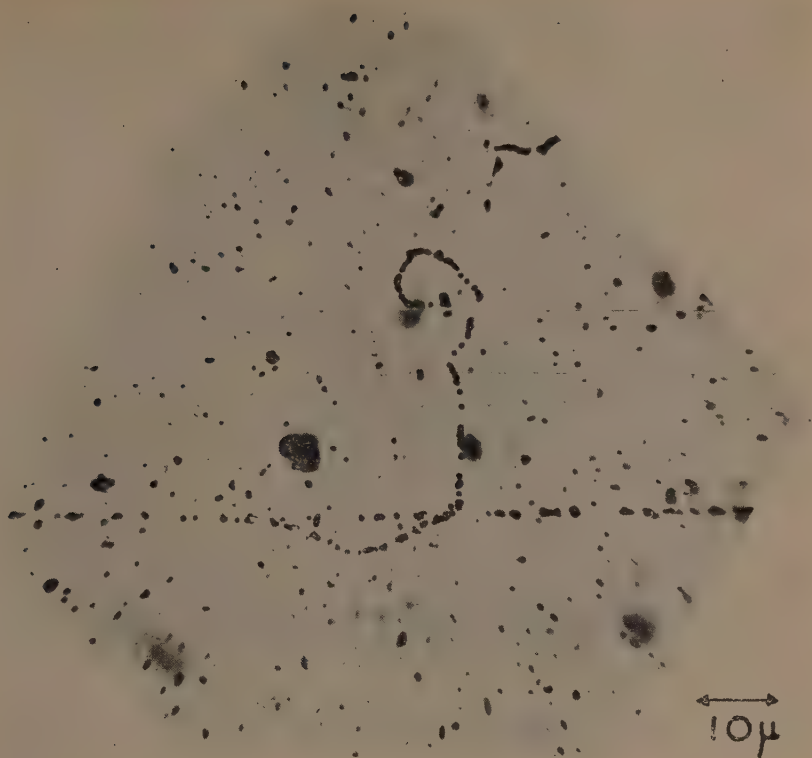
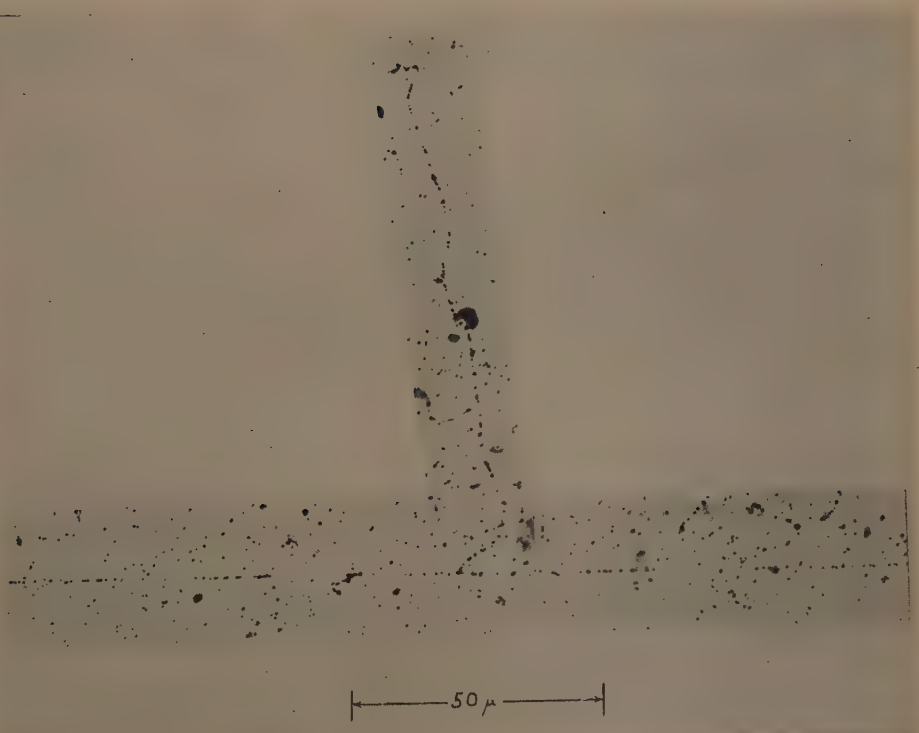
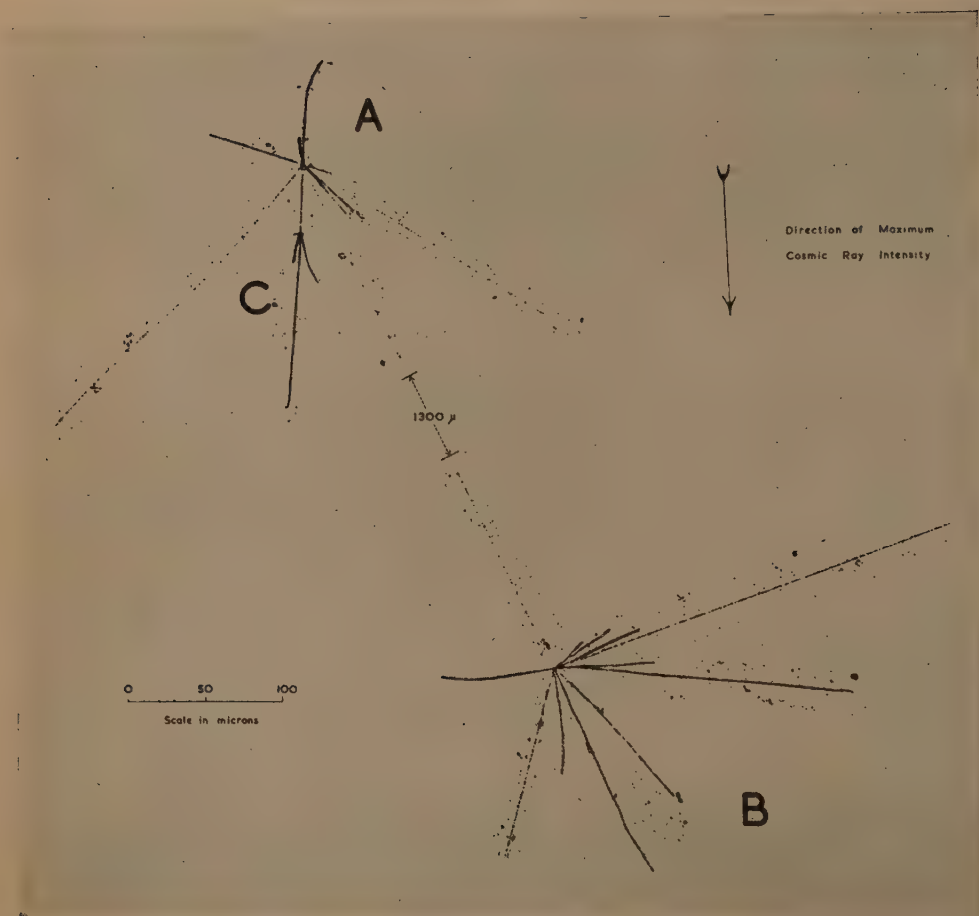


FIG. 5.



Figs. 4 & 5 Collisions between particles at minimum ionization and electrons.
(Photomicrographs: G. P. Cooke.)

FIG. 6.



Star formation by a secondary particle of minimum ionizing power.

(Photomicrograph : G. P. Cooke and Miss C. H. Lamb,
both of the Kodak Research Laboratories.)

One of the heavily ionizing particles of star A produces a third star C of five particles. This time sequence of the formation of the stars is thought to be the most probable following an examination of the angular distribution of the particles in each star compared with the incident direction of the cosmic radiation.

The authors wish to express their gratitude to Dr. E. P. George, Birkbeck College, University of London, for permission to expose the plates in his underground laboratory; to Messrs. D. W. Fry and F. K. Goward, T.R.E. Great Malvern, for permission to use the Synchrotron machine; to Professor C. F. Powell, and Mr. D. T. King, University of Bristol, for helpful discussions and Mr. M. Roberts, University of Bristol, for the exposures on the Jungfrauoch.

This note is based on work carried out under contract to the Atomic Energy Establishment, Ministry of Supply, whose permission to publish is gratefully acknowledged.

REFERENCES.

- BERRIMAN, R. W., 1948, *Nature*, **162**, 992.
 CLAY, J., WIERSMA, J. T., and JONKER, K. H. J., 1938, *K. Akad., Amsterdam, Proc.* **41**, 7, 706.
 COSYNS, M., DILWORTH, C., OCCHIALINI, G., 1949, *Note 6, Centre de Physique Nucleaire, Universite Libre de Bruxelles.* (Janvier.)
 GOLDSCHMIDT-CLERMONT, Y., KING, D. T., MUIRHEAD, H., and RITSON, D. M., 1948, *Proc. Phys. Soc.*, **61**, 2, 183.
 GREISEN, K., 1942, *Phys. Rev.* **61**, 212.
 KOLHÖRSTER, W., and JÁNOSSY, L., 1932, *Zeit. f. Phys.*, **93**, 1-2, 111.

C. Notices of New Books and Periodicals received.

A Survey of General and Applied Rheology. By G. W. SCOTT BLAIR. Second Edition. (Pitman.) Price 40s.

DR. SCOTT BLAIR has combined in this volume two related studies. The survey itself defines the field of rheology (ranging from the history of early Indian rheological conceptions to a theoretical study of baldness and practical guidance on bringing the bull to the cow), and describes the experimental methods and modern theories: this section has a full bibliography and useful summaries of papers. With this is presented a most interesting discussion of the author's own researches on "psycho-physics", in which he relates the subjective impression of "firmness" to the material parameters measured by ordinary physical methods. The exposition in the mathematical portions is obscure, and sometimes, as in the discussion of "pure" and "simple" shears, wrong, but the book will still be of great value to the discriminating reader.

F. R. N. N.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

CI. *Magnetic Properties of Nickel-Cobalt and Related Alloys*

By E. P. WOHLFARTH, Ph.D. *,

Department of Physics, Leeds University †.

[Received July 11, 1949.]

SUMMARY.

A discussion is given, on the basis of the collective electron theoretical treatment of ferromagnetism, of the magnetic properties of nickel-cobalt and related alloys. In §1 previous work and the main underlying assumptions are briefly discussed in relation to the present work. Experimental results for nickel-cobalt and nickel-iron alloys, shown in §2 in a number of figures, include the variation of the saturation moment, σ_0 , and Curie temperature, θ , with concentration, c , and of the susceptibility and spontaneous magnetization with temperature. In fig. 5 values of θ derived theoretically (from the variation of σ_0 with c) are compared with observed values, and good general agreement is found, although there are significant deviations, more pronounced for nickel-iron. Reduced magnetization, temperature curves for nickel-cobalt obey the "law of corresponding states", but those for nickel-copper do not; they tend to approach the collective electron theoretical curve as the copper concentration increases (fig. 7).

The results of the analysis are discussed qualitatively in §3. The discrepancies between estimated and observed Curie temperatures are ascribed to deviations in the energy band shape from that presupposed. The character of the variation of the magnetization, temperature curves for nickel-cobalt and nickel-copper alloys is considered in relation to previous work. The results for nickel-copper suggest that the interchange interaction energy is proportional to the square of the relative magnetization, ζ , as long as ζ^2 does not exceed about 0.9. A brief discussion is given of the way in which the magnitude of the interchange interaction energy could be estimated theoretically.

In §4 the results for other alloys are shown and discussed. For nickel-chromium and related alloys the variation of the magnetic properties with concentration is apparently due to a filling-up of the holes in the d band of nickel by some, if not all, of the outermost electrons of the added element, as in nickel-copper. The magnetic properties of nickel-manganese alloys are more complicated, and no fully satisfying interpretation of all of them can at present be given, although there is some degree of correlation with the properties of the other alloys considered.

* Now at Department of Mathematics, Imperial College, London.

† Communicated by Professor E. C. Stoner, F.R.S.

§ 1. INTRODUCTION.

IN this paper a quantitative discussion is given, on the basis of the collective electron theoretical treatment (Stoner 1938, 1948), of the magnetic properties of nickel-cobalt and, in less detail, of related alloys, mainly those of nickel with the remaining elements of the first transition period. A large part of the paper is taken up in presenting graphically and analysing the extensive experimental data which are available. The methods of analysis used are those suggested by Stoner's treatment, and the immediate conclusions follow directly from the general co-ordinating scheme. Some of the conclusions are, however, of a semi-formal character, and their physical significance requires further discussion which must, of necessity, be tentative, since it involves largely unsolved general problems in the theory of the metallic state, as distinct from problems peculiar to ferromagnetics.

This paper is a continuation of two earlier ones on the properties of nickel alloys. The first of these (Wohlfarth 1949 a) deals with the magnetic and also the thermal behaviour of the alloys with copper over the complete concentration range. For this system, and for a large number of others (*cf.* §4), there is conclusive evidence of the approximate validity of Mott's suggestions (1935) as to the effect of alloying on the electron distribution. This may be described as a filling-up of the unoccupied states in the d band of nickel by the loosely bound electrons of the added element. In the second paper (Wohlfarth 1948) a discussion is given, *inter alia*, of the properties of alloys with palladium and platinum. For nickel-palladium the magnetic results may be interpreted by taking the electron distribution to be unchanged during alloying, it being clear from independent magnetic and thermal evidence that the number of holes per atom in the d band and the unoccupied energy width are both nearly the same for the two constituent metals. The nickel-cobalt alloys are different again, in that the number of holes increases during alloying. An analysis of the experimental data indicates, in addition (*cf.* §2), that here the relative magnetization (the ratio of the number of parallel spins to the total number) at absolute zero, ζ_0 , remains constant over the whole range of cobalt concentrations. For most other nickel alloy systems, in contrast, ζ_0 decreases regularly as the concentration of the solute metal increases.

In the previous papers the simplifying assumptions on which application of the collective electron treatment to alloys is based were discussed at some length, and it will suffice here merely to recall the relevant essentials. For a quantitative application of the results of Stoner's calculations it is necessary to take the energy density of states near the Fermi limit of the hole distribution in the d band to be of the same form as for perfectly free electrons, although the holes may have a much larger effective mass. This assumption is justifiable, both theoretically and from a consideration of experimental results, for nickel and those of its alloys for which the number of holes per atom remains relatively small.

For nickel-cobalt and nickel-iron alloys, however, the number of holes becomes much larger, and marked deviations may be expected between the experimental results and those calculated on the assumption of a parabolic energy distribution of states. It is possible, in principle, to derive from such deviations information about the band structure, as will be discussed further in §§2 and 3.

A further assumption that has to be made concerns the dependence on magnetization of the total interchange interaction energy. A preliminary discussion of these and related problems has already been given (Wohlfarth 1949 b). It was suggested that Stoner's assumption of a variation of the interaction energy as the square of the magnetization is valid, in ordinary fields, at temperatures above the Curie point, but that further magnetization dependent terms may have to be taken into account below. The agreement with theory of the susceptibility results for nickel was contrasted with the marked deviations between the observed and the calculated magnetization, temperature curves. Further evidence in favour of these ideas may be obtained by considering the magnetization, temperature curves for nickel-cobalt and nickel-copper alloys for which extensive experimental data are available (§2). By considering the susceptibility results, in addition, it is possible to deduce the dependence on the concentration of the solute metal of $k\theta'/\epsilon_0$, where $k\theta'$ is Stoner's interchange interaction coefficient (*cf.* (3.2)), and ϵ_0 the unoccupied energy width of the d band. For an adequate theoretical discussion of the effects of alloying on the interchange interaction energy it is necessary to consider problems similar to those relating to the determination of the energy bands of the alloys. By developing the arguments of the previous paper (*cf.* §3) it is possible, however, without entering into details, to indicate the main controlling factors.

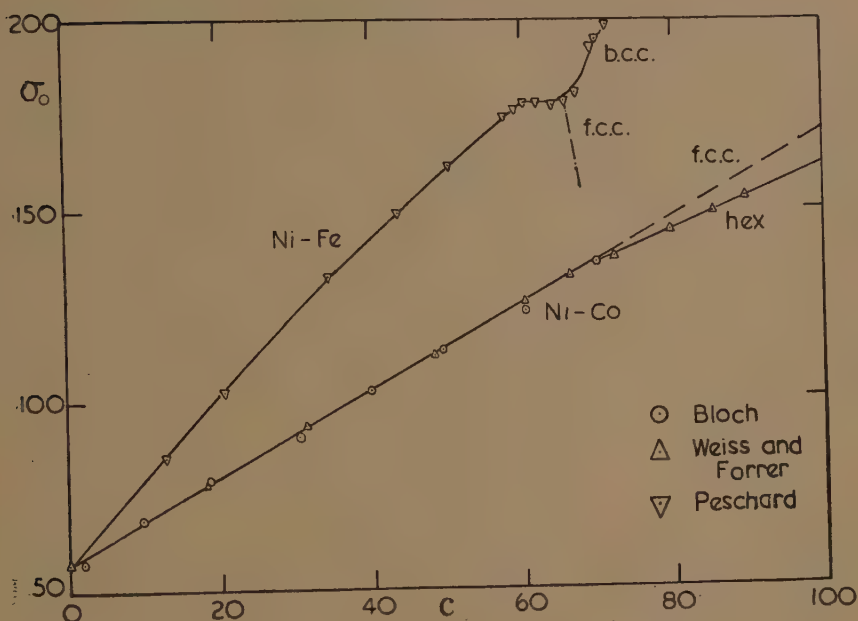
The relevant experimental data for nickel-cobalt and nickel-iron alloys, and the methods used for analysing them, are given in §2, and a qualitative discussion of the results of the analysis in §3. In §4 a brief account is given of the magnetic properties of other nickel alloys in the light of the earlier discussion.

§2. EXPERIMENTAL RESULTS.

Experimental data required include the following:—

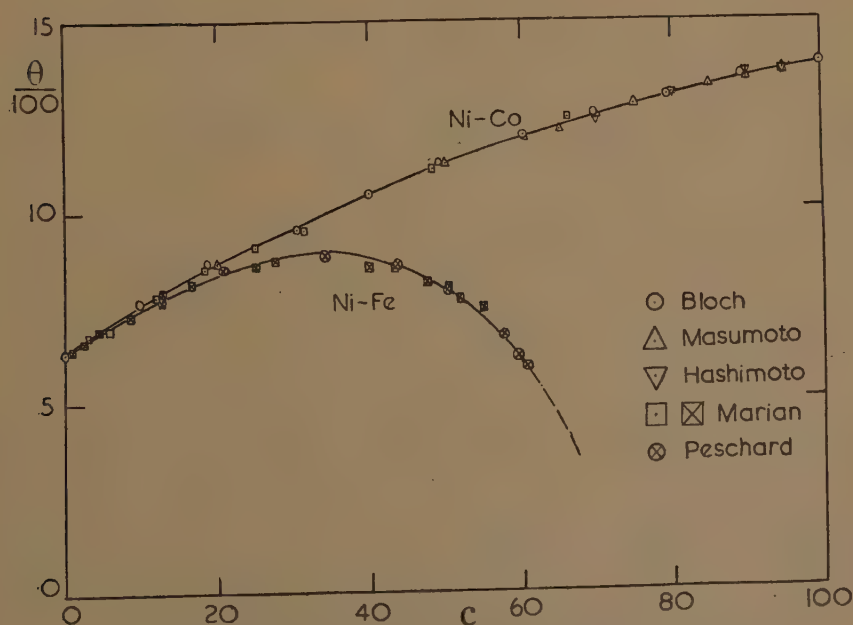
- (a) Variation of the saturation moment at absolute zero, σ_0 , with the atomic concentration of the solute metal, c .
- (b) Variation of the ferromagnetic Curie temperature, θ , with c .
- (c) Curves relating the inverse of the mass susceptibility, χ , with temperature, above θ , for a range of c values.
- (d) Curves relating the spontaneous magnetization, σ , with temperature, below θ , for a range of c values.

Fig. 1.



Saturation moment of nickel-cobalt and nickel-iron alloys.
 c , atomic concentration, per cent.

Fig. 2.

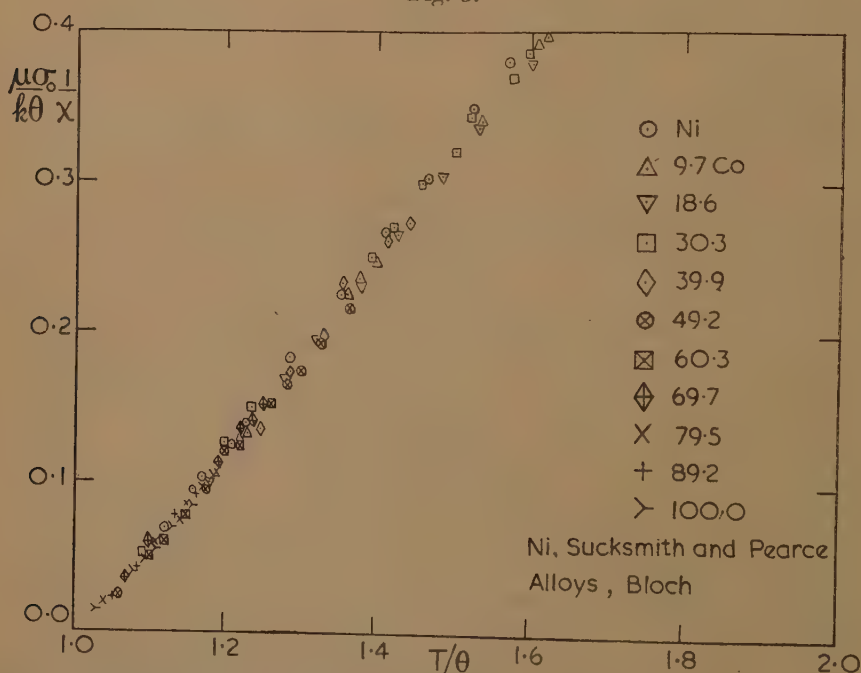


Curie temperature of nickel-cobalt and nickel-iron alloys.
 c , atomic concentration, per cent.

Experimental values of the Curie temperature for the cobalt alloys are due to Bloch (1911, 1912), Masumoto (1926), Hashimoto (1932) and Marian (unpublished, results quoted by Forrer 1933), and for the iron alloys to Peschard (1925) and Marian (*loc. cit.*). They are shown in fig. 2. For nickel-cobalt the alloys are cubic at the Curie point for all values of c , and θ is seen to increase regularly. For nickel-iron θ first increases and then decreases, with a maximum at about 34 per cent Fe. For simplicity in fig. 2 only the reversible range is covered for the iron alloys.

Values of the susceptibility of nickel-cobalt alloys have been obtained by Bloch (1911) over the whole concentration range. The reduced experimental values are shown in fig. 3, which includes the values for pure nickel due to Sucksmith and Pearce (1938). These values, as those for the other alloy systems to be discussed, may now be considered

Fig. 3.



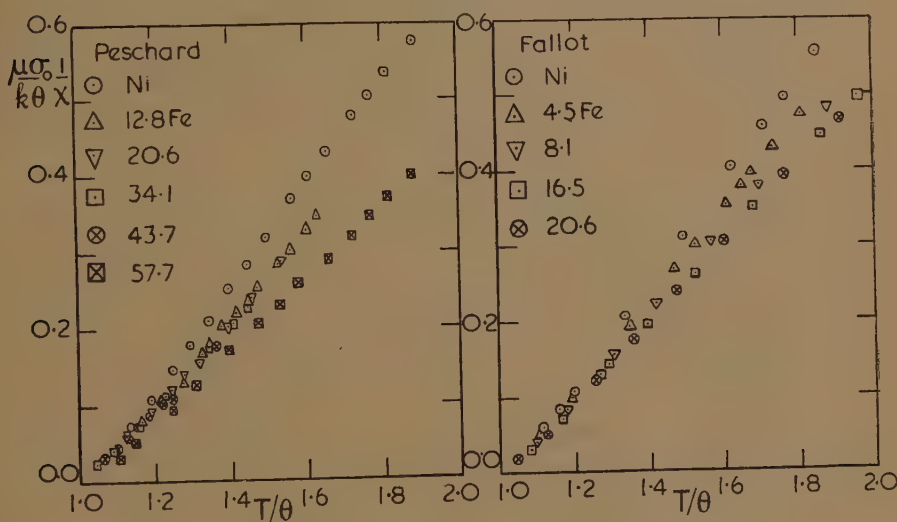
Reduced inverse mass susceptibility of nickel-cobalt alloys.
 T/θ , reduced temperature. Atomic concentration of cobalt (per cent) as indicated.

in conjunction with the collective electron theoretical curves (*cf.* Wohlfarth 1949 a, fig. 9). The results for nickel-cobalt indicate that the relative magnetization at absolute zero, ζ_0 , retains the value for nickel, previously found to be 1 (with $k\theta'/\epsilon_0$ just equal to the critical value, $2^{-1/3}$, for the attainment of saturation), over the whole range of cobalt concentrations. This is in contrast to the behaviour of all other alloy systems so far considered, including nickel-copper (*loc. cit.* figs. 10, 11 and 12) and nickel-iron (see below).

The susceptibility results for nickel-iron alloys are due to Peschard (1925) and Fallot (1944) and are shown in reduced form in fig. 4.

The two sets of results are reasonably consistent and show that ζ_0 decreases with c up to about 10 per cent Fe, then remains constant, at $\zeta_0=0.9_0$, up to about 45 per cent Fe and finally decreases again, the value for the 57.7 alloy being about 0.8_2 . Again the behaviour differs from that of other alloy systems; for nickel-copper, for example, ζ_0 decreases regularly with c , slowly for the dilute alloys and more rapidly for the alloys with a higher copper concentration (*loc. cit.*, figs. 5 and 11).

Fig. 4.



Reduced inverse mass susceptibility of nickel-iron alloys.
 T/θ , reduced temperature. Atomic concentration of iron (per cent) as indicated.

Analysis of the Results.

The results for nickel-cobalt are particularly simple, since $\zeta_0=1$ over the whole concentration range. Hence, from (2.1)

$$q = A\sigma_0/M_B; \quad \dots \quad (2.3)$$

in addition, $k\theta/\epsilon_0$ also remains constant (*cf.* (2.6)), so that, from (2.2),

$$(\theta/\theta_0) = (q/q_0)^{2/3},$$

where θ_0 , q_0 refer to pure nickel. Taking for q_0 the value obtained from (2.3), $q_0=0.60_5$,

$$\theta = (\theta_0/0.715)q^{2/3}. \quad \dots \quad (2.4)$$

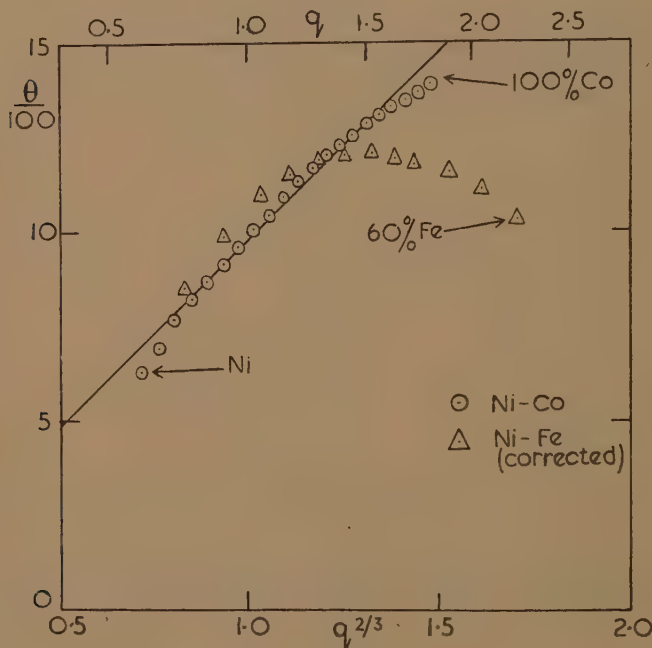
For pure (cubic) cobalt, (2.3) gives $q=1.80_5$ and for intermediate concentrations it follows from fig. 1 that

$$q = 0.60_5 + 1.20_0 (c/100), \quad \dots \quad (2.5)$$

c being the percentage concentration. Hence from (2.4) θ should increase more slowly with c than linearly, as is borne out qualitatively by the

experimental results shown in fig. 2. To obtain a quantitative comparison, θ values read off from fig. 2 (at intervals of 5 per cent in c) and plotted against $q^{2/3}$ (obtained from (2.5)) may be compared with the theoretical curve given by (2.4). The best fit over the whole range is obtained by taking $\theta_0 = 700^\circ \text{K.}$; this is rather higher than the observed Curie temperature of pure nickel (631°K.), but the discrepancy is consistent with that found for the alloys, as discussed below. The results of the analysis are shown in fig. 5, from which it appears that the observed Curie

Fig. 5.



Estimated and observed Curie temperatures of nickel-cobalt and nickel-iron alloys.

q , number of holes per atom. Theoretical curve given by

$$\theta = 0.979 \times 10^3 q^{2/3} \text{ (cf. (2.4)).}$$

(For details see text.)

temperatures vary rather more slowly with q than is given by (2.4). Since for nickel-cobalt $k\theta/\epsilon_0$ is constant, this implies that the density of states curve is rather more steep than is given by the parabolic relation (cf. (2.2)). On the whole the general agreement is, however, remarkably close; in no part of the range does the difference between the estimated and observed θ values exceed 10 per cent, the values for nickel being 700°K. and 631°K. and for cobalt 1450°K. and 1390°K. respectively.

A similar analysis for nickel-iron is much more uncertain, owing to the variation of ζ_0 (cf. fig. 4), but values of θ may be incorporated in fig. 5 as follows:—Values of $k\theta/\epsilon_0$ are calculable from those of ζ_0 using

the relations, involving Fermi-Dirac functions (Stoner 1938),

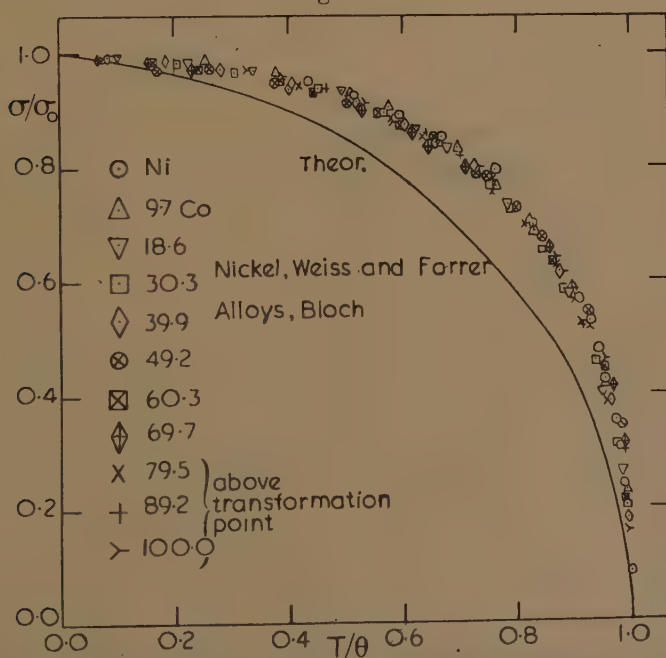
$$\left. \begin{aligned} \frac{k\theta}{\epsilon_0} \frac{F(\eta)}{F'(\eta)} = \frac{k\theta'}{\epsilon_0} = \{(1+\zeta_0)^{2/3} - (1-\zeta_0)^{2/3}\} / 2\zeta_0, \\ \text{where } F(\eta) = \frac{2}{3}(k\theta/\epsilon_0)^{-3/2}, \quad F'(\eta) = dF/d\eta. \end{aligned} \right\} \dots (2.6)$$

This gives (*loc. cit.* Table III.) $k\theta/\epsilon_0 = 0.4065$ for $\zeta_0 = 1$ ($k\theta'/\epsilon_0 = 2^{-1/3}$), and correspondingly smaller values for $\zeta_0 < 1$. The observed values of the Curie temperature for the iron alloys were multiplied by $0.4065/(k\theta/\epsilon_0)$, corresponding to the values of ζ_0 deduced from the susceptibility results, and the values "corrected" in this way are shown in fig. 5. They must be regarded as rather uncertain, mainly owing to the fact that $k\theta/\epsilon_0$ is sensitive to small changes in ζ_0 when this is near 1, but the results indicate that the discrepancies are of the same type as those found for nickel-cobalt, although they are much more pronounced, particularly for the iron-rich alloys.

Spontaneous Magnetization.

The temperature variation of spontaneous magnetization for nickel-cobalt alloys has been investigated by Bloch (1911, 1912), whose results

Fig. 6.



Reduced magnetization, temperature curves for nickel-cobalt alloys.

Atomic concentration of cobalt (per cent) as indicated.

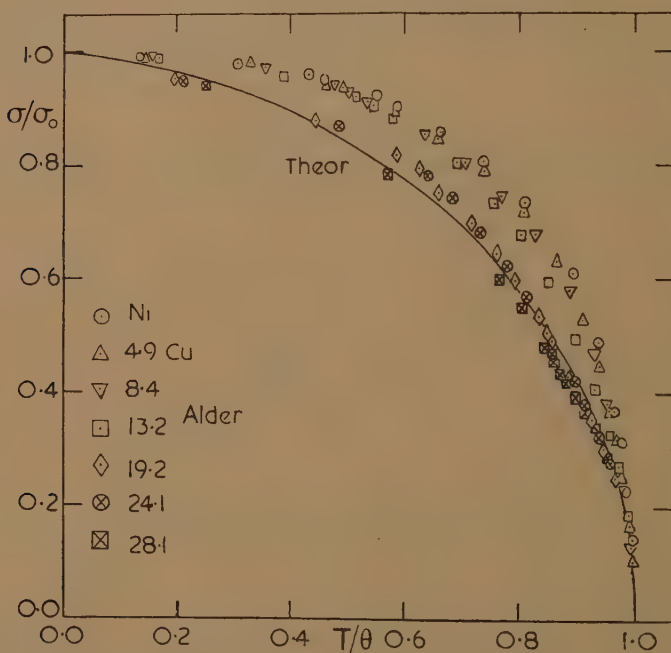
Theoretical curve due to Stoner (1938) with $\zeta_0 = 1$.

are shown in reduced form (σ/σ_0 plotted against T/θ) in fig. 6, which also includes the values for pure nickel due to Weiss and Forrer (1926; *cf.* Stoner 1936, p. 181).

It is seen that the reduced curves are practically coincident (the so-called "law of corresponding states" is obeyed). The results are, however, in disagreement with the collective electron theoretical curve, drawn in fig. 6 from Stoner's data (*loc. cit.* Table III.; $\zeta_0=1$, $k\theta'/\epsilon_0=0.794$, the value derived from the susceptibility results shown in fig. 3).

Fig. 7 gives the reduced magnetization, temperature curves for nickel-copper alloys. These are included here as they are related to the subjects discussed, and in previous papers no attention has been given, on the basis of the collective electron treatment, to the properties of this alloy system below the Curie point (*cf.* Wohlfarth 1949 a, p. 438; it was there

Fig. 7.



Reduced magnetization, temperature curves for nickel-copper alloys.

Atomic concentration of copper (per cent) as indicated.

Theoretical curve with $\zeta_0=1$, 0.9.

suggested that "the variation of spontaneous magnetization with temperature . . . would require a fuller consideration of effects . . . which lie outside the scope of the paper". Later work, described here, has, however, shown that this statement was, in the main, unfounded, and that the deviations between observed and calculated magnetization, temperature curves should be explicable by only slight extension of Stoner's ideas). The experimental values are due to Alder (1916; I am indebted to Professor W. Sucksmith, F.R.S., for the loan of a copy of Alder's *Thesis*). The results for pure nickel are in good agreement with those of Weiss and Forrer (1926, *cf.* fig. 6) and indicate the reliability

of Alder's methods of measurement. There is, however, an apparent irregularity in the variation of the properties for some of the alloys, and experiments in progress in the Physics Department, Sheffield University, are intended to test the reliability of Alder's results. Also included in fig. 7 is the theoretical curve already given in fig. 6 ($\zeta_0=1, k\theta'/\epsilon_0=0.794$; this curve is practically indistinguishable from that for $\zeta_0=0.9, k\theta'/\epsilon_0=0.733$ and thus covers the range of ζ_0 values deduced for nickel-copper alloys from the susceptibility results (*loc. cit.*). Fig. 7 shows that for nickel-copper the "law of corresponding states" no longer holds. The reduced curves for different alloys do not coincide, and it is of particular interest that with increase of copper concentration they approach the theoretical curve more and more closely.

No detailed experimental results are available for the iron alloys. In Peschard's paper (1925) results are given for only three temperatures (85, 194 and 289° K.), so that detailed analysis is impossible. The results do, however, provide definite evidence that the "law of corresponding states" is again not obeyed and that, as for nickel-copper, the reduced experimental results approach the collective electron theoretical curve as the iron concentration increases.

It has been the aim of this section to present the relevant experimental results for the alloy systems under consideration in such a form that discussion of them is immediately possible. Results for many other alloy systems have also been examined, but the data are usually much less extensive and bring out fewer novel features (an exception is nickel-manganese). It would, however, be misleading to omit reference to them entirely, and so a brief discussion is given in §4 of the magnetic properties of those nickel alloy systems which are most closely related to the ones considered in greater detail in this and the following section.

§3. DISCUSSION.

A discussion of the results of the analysis given in the preceding section must of necessity be based on the general theory of the metallic state. In spite of the many advances already made in this field, many outstanding problems remain to be solved even for pure metals. It is thus not surprising that the treatment of alloys is still in a very preliminary state, although a beginning has been made in the interpretation of some aspects of alloy behaviour by the recent work of Huang (1948) and Arafat (1949).

The simplest possible way of describing the effects of alloying on electronic structure is here adopted. Changes in the electron density are accompanied by a movement of the Fermi limit to higher or lower energies, the energy distribution of states being assumed to be of the same form as for the solvent metal. While this assumption is reasonable for alloys like nickel-copper, it is an over-simplification for the nickel-cobalt and nickel-iron alloys, for which changes in the shape, and possibly the position, of the energy bands will occur during alloying. These changes are more pronounced for nickel-iron, as seen in fig. 5, the observed

Curie temperatures differing more widely from those calculated than for nickel-cobalt. Even for the cobalt-alloys, however, the discrepancies between the observed Curie temperatures and those given by relation (2.4) indicate, as already stated, that the shape of the d band (over the range $0.6 < q < 1.8$) diverges from the parabolic shape presupposed, being apparently rather less convex than is given by the $\epsilon^{1/2}$ relation (*cf.* (2.2)). These conclusions could be confirmed most directly by considering the low temperature specific heat of the alloys. Unfortunately no experimental data are available except for pure nickel, the electronic heat of which is consistent with a Curie temperature higher than that observed (Stoner 1939, p. 368), in qualitative agreement with the conclusions reached here (*cf.* fig. 5). With the deviations from a parabolic band shape may be connected the effective decrease of the ζ_0 values of the nickel-iron alloys (fig. 4), since, for a given number of holes in the d band, an increase of the density of states at the Fermi limit over that assumed leads to an increase of the susceptibility and hence a relative decrease of the slopes of the $1/\chi$ — T curves.

A full interpretation of the differences in character of the susceptibility results requires consideration not only of the energy band structure but also of the interchange interaction energy. This problem is one of great complexity, and no more than a qualitative discussion can here be given. In the earlier paper (Wohlfarth 1949 b) it was shown, on the basis of the approximation of tight binding, that for two electrons with parallel spin and momenta \mathbf{k} , \mathbf{k}' the interaction energy contains terms of the form

$$C(\rho) \exp \{i(\mathbf{k}-\mathbf{k}') \cdot \rho\}, \quad (3.1)$$

where ρ is the nearest neighbour distance, and the coefficients $C(\rho)$ have the form of molecular (Heitler-London) integrals whose magnitude depends on the overlap of *atomic* wave functions. Since, to obtain the total energy, the terms (3.1) have to be integrated over the volume in momentum space enclosed by the Fermi surface, it is necessary to have information about the energy distribution of momentum states in this region. In addition, values of $C(\rho)$ are required, raising the whole vexed question as to which of the atoms in a ferromagnetic alloy may be regarded as in effective interaction. For what is, in this connection, the simplest case, nickel-palladium (Wohlfarth 1948), the energy distribution of states varies little during alloying, so that the changes in the magnetic properties must be ascribed principally to changes in the degree of overlap of the atomic wave functions for a pair of neighbouring unlike atoms. For the nickel-cobalt and nickel-iron alloys both the electronic structure and the amount of overlap will vary, and it is thus impossible, without much further calculation, to predict theoretically the character of the variation of ζ_0 with concentration. It is clear that the apparent simplicity of the behaviour of the cobalt alloys (fig. 3) must be largely accidental, since both of the above-mentioned factors are likely to be important.

For the susceptibility results the representation of the interchange interaction energy, J , by a single parameter, θ' , such that

$$J(\zeta) = -\frac{1}{2}Nk\theta'\zeta^2, \dots \dots \dots (3.2)$$

is justified, since ζ , the relative magnetization, is small in ordinary fields. The magnetization, temperature curves below the Curie point indicate to what extent higher terms in the expression for $J(\zeta)$ are important as $\zeta^2 \rightarrow 1$. For the nickel-cobalt alloys (figs. 3, 6), as for pure nickel, ζ varies between 0 and 1 as the temperature decreases from the Curie point to absolute zero, and the discrepancy between the observed magnetization, temperature curves and that calculated on the basis of (3.2) shows that terms in J involving higher powers of ζ than the second do, in fact, enter (alternatively, using (3.2), θ' is no longer a constant but decreases with increasing temperature below the Curie point). For the nickel-copper alloys, on the other hand, ζ_0 , the relative magnetization at absolute zero, decreases during alloying, so that ζ varies over an increasingly restricted range ($\zeta_0 > \zeta > 0$). Hence the terms in J additional to (3.2) will become more and more unimportant, and the reduced magnetization, temperature curves should agree the more closely with that calculated (using (3.2), with θ' constant) the higher the copper concentration. This is borne out by the experimental results (fig. 7), which indicate that the additional terms in J cease to be important for a copper concentration exceeding about 20 per cent. This result is to be considered in conjunction with fig. 1 of the earlier paper (Wohlfarth 1949 b) and suggests that the deviations from linearity of the $J(\zeta) - \zeta^2$ curves are only significant for values of ζ^2 exceeding about 0.9 (the value of ζ_0^2 deduced from the susceptibility results for a nickel-copper alloy with a concentration of 20 per cent). Similar considerations apply to the nickel-iron alloys (§2).

The discussion of this section has been restricted to those factors which are believed to be of primary importance in their influence on the magnetic properties of the alloys considered. Various secondary effects have been neglected. A discussion of some of these for temperatures above the Curie point has already been given (Wohlfarth 1949 a). Below the Curie point, apart from "domain effects" (Stoner 1936), the influence of thermal expansion on both band structure and interchange interaction energy may be important, as may the "transfer effect" (*loc. cit.*). In alloys the experimental difficulties involved in attaining homogeneity in composition may be considerable; fluctuations in the local intensity of magnetization may occur, as seen by the pronounced tailing of the curves in the Curie point region. These have been discussed by Néel (1940). Fluctuation effects are not negligible even for pure nickel, where a slight tailing of the magnetization, temperature curve and a more pronounced one in the specific heat curve has been observed (*cf.* Stoner 1948). A fuller discussion of these secondary effects is, however, outside the scope of the present paper.

§ 4. EXPERIMENTAL RESULTS FOR SOME RELATED ALLOYS.

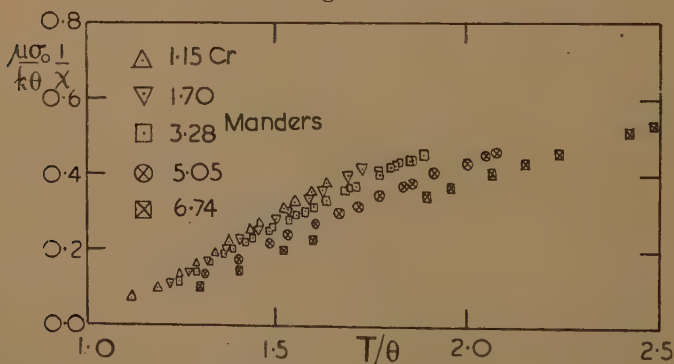
Owing to the extensive magnetic measurements of the French workers, experimental data are now available for a considerable number of alloys of nickel with non-ferromagnetic metals. Those alloys for which the solute metal has a weak, and usually constant, paramagnetism (*e.g.* nickel with gold, zinc, aluminium, antimony, silicon and tin) have magnetic properties very similar to those of the nickel-copper alloys (Stoner 1948,

Fig. 8.



Saturation moment and Curie temperature of nickel-chromium alloys.
 c , atomic concentration, per cent.

Fig. 9.



Reduced inverse mass susceptibility of nickel-chromium alloys.
 T/θ , reduced temperature.

Atomic concentration of chromium (per cent) as indicated.

Wohlfarth 1949 a). It is surprising, however, that even for those alloys for which the added elements are transition metals, the magnetic results rarely show any novel features, as exemplified by the results for nickel-chromium shown in figs. 8 and 9. The Curie temperatures and saturation moments have been measured by Sadron (1932) and Marian (1937) and the susceptibilities by Manders (1936).

The results are very similar to those for nickel-copper (*loc. cit.*), and comparison shows that chromium behaves as if it contained about four loosely bound electrons per atom which enter the d band of nickel during alloying. It would be premature to speculate more closely, on the basis of these results alone, about the band structure of chromium and, in particular, about the distribution of its six outermost electrons among the available energy bands. Further evidence that the distribution of the outermost electrons of chromium is similar to that of copper may, however, be derived from the paramagnetic susceptibility of the metal (Bates and Baqi 1936, Söchtig 1940), which is found to be relatively small (having regard to the relatively large number of electrons per atom) and to vary little with temperature. This suggests that the energy density of the states occupied by the electrons is smaller and the characteristic temperature considerably higher than for the holes in the d band of nickel.

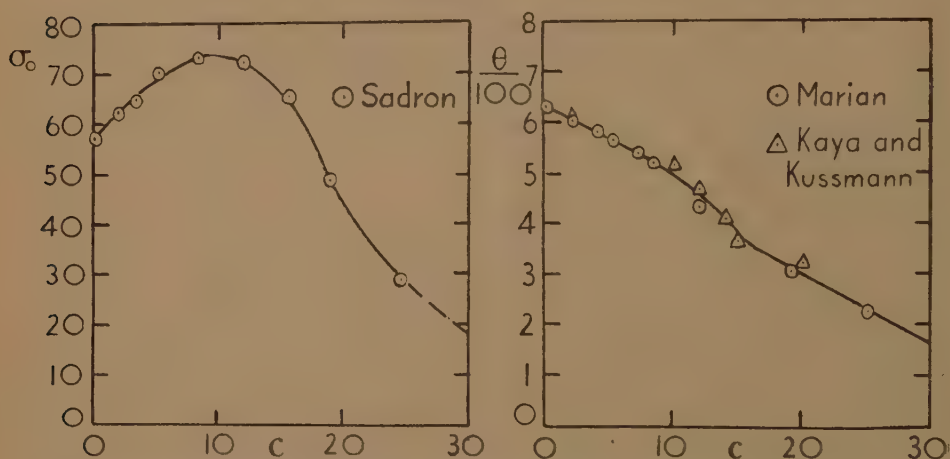
The results for nickel-vanadium and nickel-titanium alloys are similar in character to those for nickel-chromium. For nickel-vanadium the Curie temperature, concentration and saturation moment, concentration curves extrapolate to zero at about 12 per cent V, so that roughly all the five outermost electrons seem to enter the d band of nickel. The same applies to nickel-titanium. The paramagnetic susceptibilities of the metals (Klemm 1939, Squire and Kaufmann 1941) are again relatively small; for vanadium χ is independent of temperature and for titanium it increases, indicating a strong overlap of the energy bands. Among nickel alloys with elements in the second and third transition periods, nickel-molybdenum and nickel-tungsten have been investigated. The results are again similar to those for nickel-chromium, but here all the six outermost electrons seem to enter the nickel d band, the σ_0 - c and θ - c curves extrapolating to zero at about 10 per cent Mo or W. The results for nickel-ruthenium, although similar in most respects to those of the other alloys considered, show some similarity with the nickel-iron alloys as regards the susceptibility results, which indicate a more rapid initial decrease of ζ_0 than for the other alloys.

To sum up, the magnetic properties of alloys of nickel with a variety of transition metals indicate that during alloying the changes in the electronic structure are remarkably similar to those taking place in nickel-copper alloys, in that a number, if not all, of the outermost electrons of the added element enter the d band of nickel.

In view of the sharp distinction between the magnetic properties of the nickel-cobalt and nickel-iron alloys on the one hand and of the alloys considered above on the other, it is not surprising that the magnetic properties of nickel-manganese alloys are of exceptional complexity. The experimental results (due to Sadron, Marian and Manders (*loc. cit.*) and Kaya and Kussmann (1931)) are shown in figs. 10 and 11 (the remarkable changes taking place for alloys with more than about 25 per cent Mn are not shown).

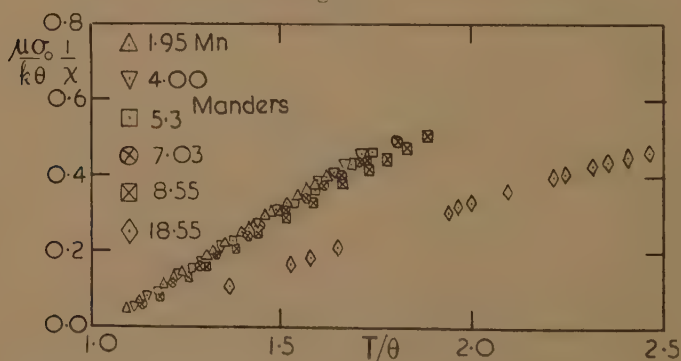
It is seen that the susceptibility results (fig. 11) are again similar to those for nickel-chromium (and nickel-copper). On the other hand, the saturation moment increases initially with manganese concentration (fig. 10), while the Curie temperature decreases (linearly except for a pronounced break at about 10 per cent Mn). In this respect there is some similarity with the nickel-iron alloys (figs. 1 and 2) for which, over part of the range, σ_0 increases while θ decreases.

Fig. 10.



Saturation moment and Curie temperature of nickel-manganese alloys.
 c , atomic concentration, per cent.

Fig. 11.



Reduced inverse mass susceptibility of nickel-manganese alloys.
 T/θ , reduced temperature.
 Atomic concentration of manganese (per cent) as indicated.

It is difficult at present to interpret the experimental results for the nickel-manganese alloys and for many other alloys to which no reference at all has been made, particularly those of iron with non-ferromagnetic metals. For the alloy systems discussed above at least a

partial solution of some of the essential problems has been obtained, and the gradual elucidation of the clearly recognized outstanding problems for these "simpler" alloys may prepare for an attack on those alloys whose detailed properties are still almost completely unexplained.

I am grateful to Professor E. C. Stoner for his many helpful suggestions and his encouragement in this and other work; to Mr. P. Rhodes for assistance in preparing this paper; and to the Department of Scientific and Industrial Research for a grant.

REFERENCES.

- ALDER, M., 1916, *Thesis*, Zürich.
 ARAFA, M. K. I., 1949, *Proc. Phys. Soc. B*, **62**, 238.
 BATES, L. F., and BAQI, A., 1936, *Proc. Phys. Soc.*, **48**, 781.
 BLOCH, O., 1911, *Vierteljahrsschr. Naturf. Ges., Zürich*, **55**, 415; 1912, *Ann. Chim. Phys., Paris*, **26**, 5.
 FALLOT, M., 1944, *J. de Phys.*, **5**, 153.
 FORRER, R., 1933, *J. de Phys.*, **4**, 501.
 HANSEN, M., 1936, *Der Aufbau der Zweistofflegierungen* (Berlin: Springer).
 HASHIMOTO, U., 1932, *Kinzoku no Kenkyu*, **9**, 57.
 HUANG, K., 1948, *Proc. Phys. Soc.*, **60**, 161.
 KAYA, S., and KUSSMANN, A., 1931, *Z. Phys.*, **72**, 293.
 KLEMM, L., 1939, *Z. Elektrochem.*, **45**, 354.
 MANDERS, C., 1936, *Ann. Phys., Paris*, **5**, 167.
 MARIAN, V., 1937, *Ann. Phys., Paris*, **7**, 459.
 MASUMOTO, H., 1926, *Sci. Rep. Tôhoku Imp. Univ.*, **15**, 463.
 MOTT, N. F., 1935, *Proc. Phys. Soc.*, **47**, 571.
 NÉEL, L., 1940, *Le magnétisme*, **2**, 65. (Paris: Institut International de Coopération Intellectuelle.)
 PESCHARD, M., 1925, *Revue de Mét.*, **22**, 490, 581, 663.
 SADRON, C., 1932, *Ann. Phys., Paris*, **17**, 371.
 SÖCHTIG, H., 1940, *Ann. Phys., Lpz.*, **38**, 97.
 SQUIRE, C. F., and KAUFMANN, A. R., 1941, *J. Chem. Phys.*, **9**, 673.
 STONER, E. C., 1936, *Phil. Trans. Roy. Soc. A.*, **235**, 165; 1938, *Proc. Roy. Soc. A*, **165**, 372; 1939, *Ibid.*, **169**, 339; 1948, *Phys. Soc. Rep. Progr. Phys.*, **11**, 43.
 SUCKSMITH, W., and PEARCE, R. R., 1938, *Proc. Roy. Soc. A*, **167**, 189.
 WEISS, P., and FORRER, R., 1926, *Ann. Phys., Paris*, **5**, 153; 1929, *Ibid.*, **12**, 279.
 WOHLFARTH, E. P., 1948, *Proc. Leeds Phil. Soc.*, **5**, 89; 1949 a, *Proc. Roy. Soc. A*, **195**, 434; 1949 b, *Phil. Mag.*, **40**, 703.

CII. *Vacuum Polarization in the Positron Theory.*

By S. T. MA,

Dublin Institute for Advanced Studies; Dublin*.

[Received July 26, 1949.]

§ 1. INTRODUCTION.

THE relativistic formalism of quantum electrodynamics of Tomonaga and Schwinger (Tomonaga 1946, Koba, Tati and Tomonaga 1947, Tati and Tomonaga 1948, Schwinger 1948 a, b, 1949) provides a new formulation of the positron theory of vacuum polarization developed by Dirac (1934), Heisenberg (1934), Serber (1935), Uehling (1935), Pauli and Rose (1936), and Weisskopf (1936). According to the Tomonaga-Schwinger formalism the induced current due to the polarization of the vacuum by an electromagnetic field is expressible in terms of the electromagnetic potential by means of an induction tensor, so that the problem of vacuum polarization reduces to the evaluation of the induction tensor.

The induction tensor has been investigated by Tati and Tomonaga, and more completely by Schwinger, Wentzel (1948), and Pauli (1949). The work of Schwinger re-establishes the previous discovery that the effect of vacuum polarization involves a divergent integral which can be interpreted as a correction to the charge of the electron and which can be eliminated by renormalizing the electronic charge. A new development in the theory is the disagreement in the views of these authors concerning the gauge invariance and the self energy of the photon, which will be discussed in some detail in the next section. This disagreement arises from the fact that the induction tensor involves certain singular functions introduced by Dirac into the positron theory. In dealing with these singular functions, Tati and Tomonaga use the well-known Fourier integral representation of these functions, while Schwinger has introduced a new integral representation of his own. It has been pointed out by Wentzel, Pauli, Tati and Tomonaga that there is great ambiguity in the result of evaluating the induction tensor owing to the presence of these singular functions and the divergent integrals, because the result depends very much on the integral representation chosen for the singular functions and on the way the divergent integrals are handled. In order to remove the ambiguity, Pauli has developed a new limiting process—the regularization—which is a more systematic and successful formulation of previous proposals of Rivier and Stueckelberg (1948) and of Feynman (1948). Pauli and Villars (1949) have shown that the induction tensor can be regularized in a mathematically rigorous way so that the gauge invariance

* Communicated by Professor W. Heitler, F.R.S.

and the vanishing of the photon self energy are ensured and the self charge of the electron is also eliminated. The latter result is also welcome because of the identity of all elementary charges.

The work carried out in the present paper may be summarized as follows: In §3 we shall give an alternative evaluation of the induction tensor along the lines of the ordinary form of perturbation theory. Our calculation will be similar to, but more in detail than, that of Tati and Tomonaga, and it differs from that of Schwinger and Wentzel in that we use the ordinary triple integral representation of the singular functions rather than Schwinger's integral representation, and that the tensor components are worked out individually rather than collectively. This work was started in an attempt to check the calculations of Schwinger and Wentzel, but it was later realized that different methods of handling the divergent integrals could lead to quite different results. The interesting feature of our rather lengthy calculation is that the divergent integrals that arise from our calculation turn out to be exactly the same as those of the counter terms introduced by Koba and Tomonaga (1948), Koba and Takeda (1948 a, b) in connection with their subtraction method, so that our calculation throws some light on the connection between the counter terms of these authors with the Tomonaga-Schwinger relativistic formalism.

Schwinger (1948 a, 1949) has investigated on the basis of the many-time formalism the effect of vacuum polarization both for the case of an external electromagnetic field and for the case of the field of a radiating electron interacting with the vacuum electrons. In §4 we shall repeat Schwinger's calculation on the basis of the usual form of perturbation theory for the energy-momentum representation of quantum electrodynamics. It will be shown that the induction tensor is the same as the result derived in §3.

In §5 we shall apply Pauli's regularization to our expression for the induction tensor. It will be shown that the regularization of the induction tensor leads to a divergence-free, relativistically covariant and gauge invariant result, which is *independent of the way the integrals are dealt with during the evaluation.*

§2. SUMMARY OF PREVIOUS RESULTS.

According to the Tomonaga-Schwinger formalism worked out to the second order in the electronic charge, the polarization of the vacuum is described by the induction tensor*

$$K_{\mu\nu}(x) = -\frac{1}{2} \left\{ \frac{\partial \Delta}{\partial x_\mu} \frac{\partial \Delta^{(1)}}{\partial x_\nu} + \frac{\partial \Delta}{\partial x_\nu} \frac{\partial \Delta^{(1)}}{\partial x_\mu} - \delta_{\mu\nu} \left(\frac{\partial \Delta}{\partial x_\lambda} \frac{\partial \Delta^{(1)}}{\partial x_\lambda} + m^2 \Delta \Delta^{(1)} \right) \right\} \epsilon(x_0). \quad (1)$$

* Strictly speaking the term "tensor" is justified only after the singularity of $K_{\mu\nu}$, which has ambiguous non-covariant contributions, is removed by subtraction or regularization. For convenience in reference, however, we shall refer to all the expressions for $K_{\mu\nu}$ in this paper as the induction tensor.

Here m is the mechanical mass of the electron, $\epsilon(x_0)=x_0/|x_0|$, and the Δ -functions are the singular functions defined by

$$-\frac{1}{2}\Delta(x)\epsilon(x_0)=\bar{\Delta}(x)=(2\pi)^{-4}\int\exp(iq_\mu x_\mu)(q_\lambda^2+m^2)^{-1}d^4q. \quad (2)$$

$$\Delta^{(1)}(x)=(2\pi)^{-3}\int\exp(iq_\mu x_\mu)\delta(q_\lambda^2+m^2)d^4q. \quad (3)$$

The following alternative forms of (1) have been given by Schwinger, Tati and Tomonaga :

$$K_{\mu\nu}(x)=\frac{\partial\bar{\Delta}}{\partial x_\mu}\frac{\partial\Delta^{(1)}}{\partial x_\nu}+\frac{\partial\bar{\Delta}}{\partial x_\nu}\frac{\partial\Delta^{(1)}}{\partial x_\mu}-\delta_{\mu\nu}\left(\frac{\partial\bar{\Delta}}{\partial x_\lambda}\frac{\partial\Delta^{(1)}}{\partial x_\lambda}+m^2\bar{\Delta}\Delta^{(1)}\right), \quad (4)$$

$$K_{\mu\nu}(x)=-i\left\{\frac{\partial\Delta^{(+)}}{\partial x_\mu}\frac{\partial\Delta^{(-)}}{\partial x_\nu}-\frac{\partial\Delta^{(-)}}{\partial x_\mu}\frac{\partial\Delta^{(+)}}{\partial x_\nu}-\frac{1}{2}\delta_{\mu\nu}\left(\frac{\partial\Delta^{(+)}}{\partial x_\lambda}\frac{\partial\Delta^{(-)}}{\partial x_\lambda}-\frac{\partial\Delta^{(-)}}{\partial x_\lambda}\frac{\partial\Delta^{(+)}}{\partial x_\lambda}+m^2\Delta^{(+)^2}-m^2\Delta^{(-)^2}\right)\right\}\epsilon(x_0). \quad (5)$$

In (5) we have

$$\Delta^{(\pm)}=\pm(\Delta^{(1)}\pm i\Delta)/2i \quad (6)$$

or

$$\Delta^{(\pm)}(x)=\pm\frac{1}{2i(2\pi)^3}\int\exp(\pm iq_\mu x_\mu)E(q)^{-1}d^3q, \quad (7)$$

where

$$E(q)=(m^2+\mathbf{q}^2)^{\frac{1}{2}}. \quad (8)$$

In deriving the Fourier integral for $\Delta^{(-)}(x)$ we have changed the sign of q_λ in (2), (3). The Fourier transform of $K_{\mu\nu}(x)$,

$$K_{\mu\nu}(p)=\int K_{\mu\nu}(x)\exp(-ip_\lambda x_\lambda)d^4x \quad (9)$$

is, as shown by Pauli (1949),

$$K_{\mu\nu}(p)=(2\pi)^{-3}\int\frac{\delta(q_\lambda^2+m^2)}{(p_\lambda-q_\lambda)^2+m^2}\{(q_\mu-p_\mu)q_\nu+(q_\nu-p_\nu)q_\mu-\delta_{\mu\nu}[(q_\lambda-p_\lambda)q_\lambda+m^2]\}d^4q. \quad (10)$$

Schwinger (1949) has shown that the functions $\bar{\Delta}(x)$ and $\Delta^{(1)}(x)$ are scalar functions of the invariant quantity $\lambda=-x_\mu^2$ and that equation (4) can be written in the form

$$K_{\mu\nu}(x)=\frac{\partial^2 G(x)}{\partial x_\mu \partial x_\nu}-\delta_{\mu\nu}H(x), \quad (11)$$

where $G(x)$ is the function of λ defined by the equation

$$\frac{d^2 G}{d\lambda^2}=2\frac{d\bar{\Delta}}{d\lambda}\frac{d\Delta^{(1)}}{d\lambda} \quad (12)$$

and the condition that it vanishes at infinity, and H is the function of λ defined by

$$\begin{aligned} H(x) &= \frac{1}{2}\square G(x)+2\frac{dG(x)}{d\lambda}+m^2\bar{\Delta}\Delta^{(1)} \\ &= -2\lambda\frac{d^2 G(x)}{d\lambda^2}-2\frac{dG(x)}{d\lambda}+m^2\bar{\Delta}\Delta^{(1)}. \end{aligned} \quad (13)$$

For the discussion of the question of gauge invariance of the theory it is convenient to rewrite (11) in the form

$$K_{\mu\nu}(x) = \left(\frac{\partial^2}{\partial x_\mu \partial x_\nu} - \delta_{\mu\nu} \square \right) G(x) - \delta_{\mu\nu} F(x), \quad . \quad . \quad . \quad (14)$$

where

$$F(x) = H(x) - \square G(x) \quad . \quad . \quad . \quad . \quad . \quad (15)$$

or

$$F(x) = 2\lambda \frac{d^2 G(x)}{d\lambda^2} + 6 \frac{dG(x)}{d\lambda} + m^2 \bar{\Delta} \Delta^{(1)}. \quad . \quad . \quad . \quad (16)$$

The Fourier transform of equation (14) is

$$K_{\mu\nu}(p) = (\delta_{\mu\nu} p_\lambda^2 - p_\mu p_\nu) G(p) - \delta_{\mu\nu} F(p). \quad . \quad . \quad . \quad (17)$$

As shown by Schwinger (1948 a, 1949) the induced current $\delta j_\mu(x)$ due to the polarization of the vacuum by an electromagnetic potential $A_\mu(x)$ is, to the second order in e ,

$$\delta j_\mu(x) = -4e^2 \int K_{\mu\nu}(x-x') A_\nu(x') d^4x' \quad . \quad . \quad . \quad (18)$$

which may also be written in the form

$$\delta j_\mu(p) = -4e^2 K_{\mu\nu}(p) A_\nu(p). \quad . \quad . \quad . \quad (19)$$

It follows from equations (14), (17), (18), (19) that the G -part of $K_{\mu\nu}$ gives rise to an induced current δj_μ which depends only on the original current j_μ and not on the potential A_μ . On the other hand, the F -part of $K_{\mu\nu}$ generally gives rise to a gauge-dependent induced current. This can be seen by reference to the Maxwell equations

$$\left(\frac{\partial^2}{\partial x_\mu \partial x_\nu} - \delta_{\mu\nu} \square \right) A_\nu(x) = 4\pi j_\mu(x), \quad . \quad . \quad . \quad (20)$$

$$(\delta_{\mu\nu} p_\lambda^2 - p_\mu p_\nu) A_\nu(p) = 4\pi j_\mu(p), \quad . \quad . \quad . \quad (21)$$

or the condition for gauge invariance given by Pauli (1949)

$$\frac{\partial K_{\mu\nu}(x)}{\partial x_\nu} = 0, \quad . \quad . \quad . \quad (22)$$

$$K_{\mu\nu}(p) p_\nu = 0. \quad . \quad . \quad . \quad (23)$$

A question closely connected with that of the gauge invariance is that of the photon self energy. Wentzel has shown that the self energy of the photon due to the creation of virtual electron pairs is given by

$$\Delta E_{\text{photon}} = 8\pi e^2 K_{11}(p) / |\mathbf{p}|, \quad . \quad . \quad . \quad (24)$$

where p_λ denotes the energy and momentum of the photon, the index 1 denotes the direction of polarization of the photon and the units are chosen such that $c = \hbar = 1$, $e^2 = 1/137$. In the case when the function $G(p)$ is non-singular on the cone $p_\lambda^2 = 0$, the G -part of $K_{\mu\nu}$ contributes nothing to the self energy of the photon and (24) reduces to

$$\Delta E_{\text{photon}} = -8\pi e^2 [F(p) / |\mathbf{p}|]_{p_\lambda^2=0}. \quad . \quad . \quad . \quad (25)$$

The deductions that follow from the above general theory have been worked out by several authors. Tati and Tomonaga have shown that (5) and (7) lead to an infinite level shift of the vacuum and an infinite self energy of a photon. These authors have also pointed out the ambiguity of their result.

Schwinger has evaluated the function $G(p)$ by means of his integral representation of the Δ -functions. The result is

$$G(p)=G'+(4\pi)^{-2}f(p), \quad . \quad . \quad . \quad . \quad . \quad (26)$$

where G' is a logarithmically divergent integral that can be eliminated by renormalizing the electronic charge and f is the invariant function of p_λ

$$f(p)=-p_\lambda^2\int_0^1\frac{z^2-z^4/3}{4m^2+p_\lambda^2(1-z^2)}dz, \quad . \quad . \quad . \quad . \quad (27)$$

first obtained by Pauli and Rose (1936). It is the integral $f(p)$ that describes the physically significant effects of vacuum polarization. From a general consideration concerning $K_{\mu\nu}(x)$ Schwinger has concluded that the function F vanishes and that the theory is gauge invariant and does not give rise to a photon self energy. This conclusion has not been corroborated by direct calculations. Wentzel has shown that the self energy given by (24), (9), (4) is either infinitely large or, if Schwinger's integral representation is used for the Δ -functions, finite and non-vanishing. The latter result can also be obtained from some more general calculations of Pauli about the expression (10) for $K_{\mu\nu}(p)$. It is easy to confirm the results of Wentzel and Pauli by evaluating the Fourier transform of the expression (16) for $F(x)$ by means of Schwinger's integral representation. The result for an arbitrary value of p_λ is

$$F(p)=- (4\pi)^{-2}(m^2+p_\lambda^2/6). \quad . \quad . \quad . \quad . \quad (28)$$

From (25) and (28) we immediately obtain Wentzel's result for the photon self energy by putting $p_\lambda^2=0$. Wentzel has pointed out that a non-vanishing photon self energy formally appearing in the theory is an undesirable feature even if it can be subtracted to fit the empirical fact. It is therefore satisfactory that the function F vanishes after the regularization is applied, so that the regularization leads in a mathematical rigorous way to gauge invariance and the vanishing of the photon self energy.

§ 3. AN ALTERNATIVE EVALUATION OF $K_{\mu\nu}(p)$.

The expression (5) for $K_{\mu\nu}(x)$ may be written in the form

$$K_{\mu\nu}(x)=A_{\mu\nu}(x)+\delta_{\mu\nu}[B(x)+m^2C(x)], \quad . \quad . \quad . \quad (29)$$

with
$$A_{\mu\nu}(x)=-i\left(\frac{\partial\Delta^{(+)}}{\partial x_\mu}\frac{\partial\Delta^{(+)}}{\partial x_\nu}-\frac{\partial\Delta^{(-)}}{\partial x_\mu}\frac{\partial\Delta^{(-)}}{\partial x_\nu}\right)\epsilon(x_0), \quad . \quad . \quad (30)$$

$$B(x)=\frac{i}{2}\left(\frac{\partial\Delta^{(+)}}{\partial x_\lambda}\frac{\partial\Delta^{(+)}}{\partial x_\lambda}-\frac{\partial\Delta^{(-)}}{\partial x_\lambda}\frac{\partial\Delta^{(-)}}{\partial x_\lambda}\right)\epsilon(x_0), \quad . \quad . \quad (31)$$

$$C(x)=\frac{i}{2}(\Delta^{(+)\,2}-\Delta^{(-)\,2})\epsilon(x_0). \quad . \quad . \quad . \quad . \quad (32)$$

The Fourier transform of equation (29) is

$$K_{\mu\nu}(p) = A_{\mu\nu}(p) + \delta_{\mu\nu}[B(p) + m^2 C(p)]. \quad (33)$$

The quantities $A_{\mu\nu}(p)$, $B(p)$, $C(p)$ can be dealt with in the same way. We give the evaluation of $C(p)$ as an example. It follows from (32), (7) that

$$C(x) = \frac{-i}{8(2\pi)^6} \iint \frac{1}{EE'} \{ \exp[i(\mathbf{q} + \mathbf{q}') \cdot \mathbf{x}] - \exp[-i(\mathbf{q} + \mathbf{q}') \cdot \mathbf{x}] \} \epsilon(x_0) d^3q d^3q', \quad (34)$$

where $E = E(q)$, $E' = E(q')$. Using the general formula

$$\int \exp(i\mathbf{l}_\lambda x_\lambda) \epsilon(x_0) d^4x = -2i(2\pi)^3 \delta(\mathbf{l}) / l_0, \quad (35)$$

we obtain

$$C(p) = \frac{-1}{4(2\pi)^3} \iint \frac{1}{EE'} \left\{ \frac{\delta(\mathbf{q} + \mathbf{q}' - \mathbf{p})}{E + E' - p_0} + \frac{\delta(\mathbf{q} + \mathbf{q}' + \mathbf{p})}{E + E' + p_0} \right\} d^3q d^3q'. \quad (36)$$

Changing the signs of \mathbf{q} and \mathbf{q}' in the second term of the integrand, we obtain

$$C(p) = \frac{-1}{2(2\pi)^3} \iint \frac{1}{EE'} \frac{E + E'}{(E + E')^2 - p_0^2} \delta(\mathbf{q} + \mathbf{q}' - \mathbf{p}) d^3q d^3q'. \quad (37)$$

Put

$$\mathbf{q} = \mathbf{p}/2 + \mathbf{k}, \quad \mathbf{q}' = \mathbf{p}/2 - \mathbf{k}. \quad (38)$$

Then we have

$$C(p) = \frac{-1}{2(2\pi)^3} \int \frac{1}{EE'} \frac{E + E'}{(E + E')^2 - p_0^2} d^3k. \quad (39)$$

Similarly,

$$A_{\mu\nu}(p) = \frac{-1}{2(2\pi)^3} \int \frac{q_\mu q'_\nu}{EE'} \frac{(E + E')(1 + a_{\mu\nu}) + p_0(1 - a_{\mu\nu})}{(E + E')^2 - p_0^2} d^3k, \quad (40)$$

and

$$B(p) = \frac{1}{2(2\pi)^3} \int \frac{q_\lambda q'_\lambda}{EE'} \frac{E + E'}{(E + E')^2 - p_0^2} d^3k. \quad (41)$$

In (40), (41), $q_\lambda = iE$, $q'_\lambda = iE'$, and the symbols $a_{\mu\nu}$ are defined as follows:

$$\left. \begin{aligned} a_{\mu\nu} &= 1 \quad \text{for } \mu, \nu = 1, 2, 3, & \text{or } \mu = \nu = 4, \\ a_{\mu\nu} &= -1 \quad \text{for } \mu = 1, 2, 3, & \nu = 4 \text{ or } \nu = 1, 2, 3, \quad \mu = 4. \end{aligned} \right\} \quad (42)$$

The expressions for $K_{\mu\nu}(p)$ corresponding to (36) and (39) respectively are

$$K_{\mu\nu}(p) = \frac{1}{4(2\pi)^3} \iint \frac{1}{EE'} [-q_\mu q'_\nu - q_\nu q'_\mu + \delta_{\mu\nu}(q_\lambda q'_\lambda - m^2)] \\ \times \left[\frac{\delta(\mathbf{q} + \mathbf{q}' - \mathbf{p})}{E + E' - p_0} + \frac{\delta(\mathbf{q} + \mathbf{q}' + \mathbf{p})}{E + E' + p_0} \right] d^3q d^3q', \quad (43)$$

$$K_{\mu\nu}(p) = \frac{-1}{2(2\pi)^3} \int \frac{1}{EE'} \frac{1}{(E + E')^2 - p_0^2} \{ q_\mu q'_\nu [(E + E')(1 + a_{\mu\nu}) + p_0(1 - a_{\mu\nu})] \\ + \delta_{\mu\nu}(E + E')(-q_\lambda q'_\lambda + m^2) \} d^3k. \quad (44)$$

The above results can also be derived from formula (10). We have, from (10),

$$A_{\mu\nu}(p) = (2\pi)^{-3} \int \frac{\delta(q_\lambda^2 + m^2)}{(p_\lambda - q_\lambda)^2 + m^2} [(q_\mu - p_\mu)q_\nu + (q_\nu - p_\nu)q_\mu] d^4q, \quad (45)$$

$$B(p) = -(2\pi)^{-3} \int \frac{\delta(q_\lambda^2 + m^2)}{(p_\lambda - q_\lambda)^2 + m^2} (q_\mu - p_\mu)q_\mu d^4q, \quad \dots \quad (46)$$

$$C(p) = -(2\pi)^{-3} \int \frac{\delta(q_\lambda^2 + m^2)}{(p_\lambda - q_\lambda)^2 + m^2} d^4q. \quad \dots \quad (47)$$

These quadruple integrals can be reduced to triple integrals by using the formula

$$\delta(q_\lambda^2 + m^2) = \delta(\mathbf{q}^2 - E^2) = [\delta(|\mathbf{q}| - E) + \delta(|\mathbf{q}| + E)]/2E. \quad \dots \quad (48)$$

For example,

$$C(p) = \frac{-1}{2(2\pi)^3} \int \frac{1}{E} \left[\frac{1}{E'^2 - (p_0 - E)^2} + \frac{1}{E'^2 - (p_0 + E)^2} \right] d^3q \quad (49)$$

or

$$C(p) = \frac{-1}{2(2\pi)^3} \int \frac{1}{EE'} \left[\frac{E' + E}{(E' + E)^2 - p_0^2} + \frac{E' - E}{(E' - E)^2 - p_0^2} \right] d^3q. \quad (50)$$

Equation (50) reduces to equation (39) if we replace the variable of integration \mathbf{q} by the variable \mathbf{k} defined by (38), and use the fact that the second term in the integrand in (50) is an odd function of \mathbf{k} .

It is important to observe that in the above we have frequently changed the sign of the 3-vectors which play the part of the variable of integration. This does not lead to any ambiguity in the case of integrals which are convergent. In the case of divergent integrals, however, the result depends very much on the way the integration is carried out. Since the reversal of the direction of a 3-vector is not a Lorentz-invariant procedure, we cannot expect our results to retain always the covariant form. This is clearly illustrated by the following example. The four integrals

$$X_\mu = \int (k_\mu/k_0) d^3k$$

with

$$k_4 = ik_0, \quad k_0 = E(k)$$

transform obviously like the components of a 4-vector. Changing the sign of \mathbf{k} , we obtain

$$X_\mu = \int (k'_\mu/k_0) d^3k,$$

where

$$k'_\mu = (-\mathbf{k}, ik_0).$$

The average values of these results are

$$X_j = 0 \quad (j=1, 2, 3), \quad X_4 = i \int d^3k,$$

which no longer transform like the components of a 4-vector. Furthermore, it should be noted that the transformation (38) also plays an important rôle in determining the final form of our result for $K_{\mu\nu}(p)$.

The triple integrals (39), (40), (41) can all be reduced by a method of Pauli and Rose (1936). Following these authors we first carry out the transformation from the variables \mathbf{k} to the variables ϕ, v, w , where ϕ is the aximuthal angle around the axis along the 3-vector \mathbf{p} , and

$$v = \frac{1}{2}(\mathbf{E} - \mathbf{E}'), \quad w = \frac{1}{2}(\mathbf{E} + \mathbf{E}'). \quad . \quad . \quad . \quad (51)$$

We next carry out the transformation from v, w to the variables y, z defined by

$$y = 2v/|\mathbf{p}|, \quad z = \left(1 - \frac{4m^2}{4w^2 - \mathbf{p}^2}\right)^{\frac{1}{2}}. \quad . \quad . \quad . \quad (52)$$

The limits of integration in the $\phi y z$ -space corresponding to the whole of the \mathbf{k} -space are

$$0 \leq \phi \leq 2\pi, \quad -z \leq y \leq z, \quad 0 \leq z \leq 1. \quad . \quad . \quad . \quad (53)$$

The elements of volume are connected by the relations

$$\frac{d^3k}{\mathbf{E}\mathbf{E}'} = \frac{2}{|\mathbf{p}|} d\phi dv dw = \frac{m^2 z}{w(1-z^2)^2} d\phi dy dz. \quad . \quad . \quad . \quad (54)$$

We first reduce the triple integrals to double integrals by performing the integration with respect to ϕ . For example, we have from (51), (52),

$$(\mathbf{E} + \mathbf{E}')^2 - p_0^2 = 4w^2 - p_0^2 = \frac{4m^2}{1-z^2} + p_\lambda^2, \quad . \quad . \quad . \quad (55)$$

so that, from (39),

$$C(p) = \frac{-m^2}{(2\pi)^2} \int_0^1 \int_{-z}^z \frac{z}{(1-z^2)[4m^2 + p_\lambda^2(1-z^2)]} dy dz. \quad . \quad . \quad (56)$$

Similarly, from (41),

$$B(p) = \frac{-m^4}{(2\pi)^2} \int_0^1 \int_{-z}^z \frac{z(1+z^2)}{(1-z^2)^2[4m^2 + p_\lambda^2(1-z^2)]} dy dz. \quad . \quad . \quad (57)$$

Substituting the results for $A_{\mu\nu}(p)$, $B(p)$ and $C(p)$ into (33), we obtain

$$K_{ij}(p) = \frac{1}{2}[(\mathbf{R} - \mathbf{T})\delta_{ij} + (3\mathbf{T} + \frac{1}{2}\mathbf{p}^2\mathbf{S} - \mathbf{R})p_i p_j] |\mathbf{p}|^{-2} \quad (i, j = 1, 2, 3), \quad . \quad (58)$$

$$K_{4\mu}(p) = K_{\mu 4}(p) = \frac{1}{4}\mathbf{S}(p_4 p_\mu - \delta_{4\mu} p_\lambda^2) \quad (\mu = 1, 2, 3, 4), \quad . \quad . \quad . \quad (59)$$

where $\mathbf{R}, \mathbf{S}, \mathbf{T}$ are the double integrals

$$\mathbf{R} = \frac{-m^4}{2\pi^2} \int_0^1 \int_{-z}^z \frac{z(3-z^2)}{(1-z^2)^2[4m^2 + p_\lambda^2(1-z^2)]} dy dz, \quad . \quad . \quad (60)$$

$$\mathbf{S} = \frac{-m^2}{2\pi^2} \int_0^1 \int_{-z}^z \frac{z(1-y^2)}{(1-z^2)[4m^2 + p_\lambda^2(1-z^2)]} dy dz, \quad . \quad . \quad (61)$$

$$\mathbf{T} = \frac{-m^4}{2\pi^2} \int_0^1 \int_{-z}^z \frac{z(1-y^2)}{(1-z^2)^2[4m^2 + p_\lambda^2(1-z^2)]} dy dz. \quad . \quad . \quad (62)$$

On performing the integration with respect to y , these double integrals reduce to

$$R=3(4\pi)^{-2}[p_\lambda^2(J+f)-4I], \quad . \quad . \quad . \quad . \quad . \quad (63),$$

$$S=-(2\pi)^{-2}(J+f), \quad . \quad . \quad . \quad . \quad . \quad (64),$$

$$T=(4\pi)^{-2}[p_\lambda^2(J+f)-4I], \quad . \quad . \quad . \quad . \quad . \quad (65),$$

where I and J are two divergent integrals independent of p_λ , namely

$$I=m^2 \int_0^1 \frac{z^2-z^4/3}{(1-z^2)^2} dz, \quad . \quad . \quad . \quad . \quad . \quad (66),$$

$$J= \int_0^1 \frac{z^2-z^4/3}{1-z^2} dz, \quad . \quad . \quad . \quad . \quad . \quad (67),$$

and $f=f(p)$ is the integral given by (27). The two integrals J_1 and J_2 in Pauli and Rose's paper are in our notation $J+f$ and J respectively. Substituting (63), (64), (65) into (58), (59), we obtain

$$K_{ij}(p)=(4\pi)^{-2}[(\delta_{ij}p_\lambda^2-p_i p_j)(J+f)-4\delta_{ij}I], \quad . \quad . \quad . \quad (68),$$

$$K_{4\mu}(p)=K_{\mu 4}(p)=(4\pi)^{-2}(\delta_{4\mu}p_\lambda^2-p_4 p_\mu)(J+f). \quad . \quad . \quad . \quad (69),$$

We note that the part of $K_{\mu\nu}(p)$ involving the integral $J+f$ is covariant, but the part of $K_{\mu\nu}(p)$ involving the integral I is not covariant.

The integrals R , S , T can be transformed back into triple integrals with respect to the variables \mathbf{k} by means of (51), (52). The result is

$$R=\frac{-1}{2(2\pi)^3} \int \frac{\mathbf{E}+\mathbf{E}'}{\mathbf{E}\mathbf{E}'} \frac{(\mathbf{E}+\mathbf{E}')^2-\mathbf{p}^2+2m^2}{(\mathbf{E}+\mathbf{E}')^2-p_0^2} d^3k, \quad . \quad . \quad . \quad . \quad (70),$$

$$S=\frac{1}{(2\pi)^3 \mathbf{p}^2} \int \frac{\mathbf{E}+\mathbf{E}'}{\mathbf{E}\mathbf{E}'} \frac{(\mathbf{E}-\mathbf{E}')^2-\mathbf{p}^2}{(\mathbf{E}+\mathbf{E}')^2-p_0^2} d^3k, \quad . \quad . \quad . \quad . \quad (71),$$

$$T=\frac{1}{4(2\pi)^3 \mathbf{p}^2} \int \frac{\mathbf{E}+\mathbf{E}'}{\mathbf{E}\mathbf{E}'} \frac{[(\mathbf{E}-\mathbf{E}')^2-\mathbf{p}^2][(\mathbf{E}+\mathbf{E}')^2-\mathbf{p}^2]}{(\mathbf{E}+\mathbf{E}')^2-p_0^2} d^3k. \quad . \quad . \quad (72),$$

Integration over all directions reduce these triple integrals to single integrals with respect to $k=|\mathbf{k}|$. In the following discussion we shall need the asymptotic behaviour of the integrands of these integrals. Expanding the integrands in descending powers of k for large values of k and retaining only the divergent parts, we find

$$R \sim \frac{-1}{2\pi^2} \int^\infty \left(k - \frac{p_\lambda^2}{4k} \right) dk, \quad . \quad . \quad . \quad . \quad (73),$$

$$S \sim \frac{-1}{6\pi^2} \int^\infty \frac{1}{k} dk, \quad . \quad . \quad . \quad . \quad (74),$$

$$T \sim \frac{-1}{6\pi^2} \int^\infty \left(k - \frac{p_\lambda^2}{4k} \right) dk. \quad . \quad . \quad . \quad . \quad (75),$$

Comparing these results with (63), (64), (65), we find

$$I \sim \frac{2}{3} \int_0^\infty k \, dk, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (76)$$

$$J \sim \frac{2}{3} \int_0^\infty \frac{1}{k} \, dk. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (77)$$

From our result (68), (69), we can draw some inference about the dependence of the induced current on the original current and its electromagnetic potential. For this purpose it is convenient to separate $K_{\mu\nu}(p)$ into a convergent part and a divergent part, namely

$$K_{\mu\nu}(p) = K_{\mu\nu}^0(p) + K'_{\mu\nu}(p) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (78)$$

with

$$K_{\mu\nu}^0(p) = (4\pi)^{-2} (\delta_{\mu\nu} p_\lambda^2 - p_\mu p_\nu) f(p) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (79)$$

and

$$K'_{ij}(p) = (4\pi)^{-2} [(\delta_{ij} p_\lambda^2 - p_i p_j) J - 4\delta_{ij} I], \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (80)$$

$$K'_{4\mu}(p) = K'_{\mu 4}(p) = (4\pi)^{-2} (\delta_{4\mu} p_\lambda^2 - p_4 p_\mu) J. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (81)$$

The two parts of the induced current corresponding to $K_{\mu\nu}^0$ and $K'_{\mu\nu}$ are given by, on account of (19), (21),

$$\delta j_\mu^0(p) = -\frac{e^2}{\pi} f(p) j_\mu(p), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (82)$$

$$\delta \mathbf{j}'(p) = -\frac{e^2}{\pi} \mathbf{J} \mathbf{j}(p) + \frac{e^2}{\pi^2} \mathbf{I} \mathbf{A}(p), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (83)$$

$$\delta j_4'(p) = -\frac{e^2}{\pi} J j_4(p). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (84)$$

Since I and J are independent of p_λ , (83), (84) can also be written in the form

$$\delta \mathbf{j}'(x) = -\frac{e^2}{\pi} \mathbf{J} \mathbf{j}(x) + \frac{e^2}{\pi^2} \mathbf{I} \mathbf{A}(x), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (85)$$

$$\delta j_4'(x) = -\frac{e^2}{\pi} J j_4(x). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (86)$$

Equation (82) gives the physically significant part of the induced current. The J -part of $\delta j_\mu'$ may be interpreted as the current due to the correction to the electronic charge. The I -part of $\delta \mathbf{j}'$ upsets the gauge covariance and also gives rise to a photon self energy which is equal to, by (24), (68),

$$\Delta E_{\text{photon}} = -2e^2 I / \pi |\mathbf{p}|. \quad . \quad . \quad . \quad . \quad . \quad . \quad (87)$$

It follows from the above discussion that a finite, relativistically covariant and gauge covariant result will be obtained if we subtract $K'_{\mu\nu}(p)$ from $K_{\mu\nu}(p)$ and $\delta j_\mu'$ from δj_μ .

Our result for $K_{\mu\nu}$ and δj_μ involves two divergent integrals, I and J , which diverge in the same way as those of the counter terms introduced by Koba and Tomonaga, Koba and Takeda. According to detailed analysis of the radiation problems by these authors, all the divergent integrals due to the effect of vacuum polarization to the second order in

e can be eliminated by including into the Hamiltonian the counter terms

$$\frac{\delta e}{e} \int \mathbf{j}(x) \cdot \mathbf{A}^T(x) d^3x - \frac{\delta e}{e} \iint \frac{j_0(x)j_0(x')}{|\mathbf{x} - \mathbf{x}'|} d^3x d^3x' + \frac{e^2}{3\pi^2} \int^\infty k dk \int \mathbf{A}^T(x)^2 d^3x, \quad (88)$$

where

$$\frac{\delta e}{e} = -\frac{e^2}{3\pi} \int^\infty \frac{1}{k} dk \quad (89)$$

and \mathbf{A}^T is the transverse part of the electromagnetic potentials. From the asymptotic formulas (76), (77) it can be seen that I and J diverge in the same way as the two divergent integrals in (88), (89). We can even obtain a closer formal connection by considering the scalar product $\delta j_\mu(x)A_\mu(x)$, where the $A_\mu(x)$ are the electromagnetic potentials in the Heisenberg picture of quantum electrodynamics and so satisfy the inhomogeneous Maxwell equation (20). Using (85), (86), we obtain for the divergent part of this product

$$\delta j'_\mu(x)A_\mu(x) = -\frac{e^2}{\pi^2} [\pi \mathbf{J} j_\mu(x)A_\mu(x) - \mathbf{I} \mathbf{A}(x)^2], \quad (90)$$

so that, by (76), (77),

$$\delta j'_\mu(x)A_\mu(x) \sim -\frac{2e^2}{3\pi^2} \left[\pi j_\mu(x)A_\mu(x) \int^\infty \frac{1}{k} dk - \mathbf{A}(x)^2 \int^\infty k dk \right]. \quad (91)$$

Hence

$$\frac{1}{2} \int \delta j_\mu(x)A_\mu(x) d^3x \sim -\frac{e^2}{3\pi} \int^\infty \frac{1}{k} dk \int j_\mu(x)A_\mu(x) d^3x + \frac{e^2}{3\pi^2} \int^\infty k dk \int \mathbf{A}(x)^2 d^3x. \quad (92)$$

Comparing (92) with (88), (89) we see that they are connected by the gauge transformation (see Heitler 1944, Appendix II.3)

$$\mathbf{A}(x) \rightarrow \mathbf{A}^T(x), \quad A_0(x) \rightarrow \int \frac{j_0(x')}{|\mathbf{x} - \mathbf{x}'|} d^3x'. \quad (93)$$

Thus our discussion throws some light on the connection between the Tomonaga-Schwinger relativistic theory and the subtraction method of counter terms. It has been pointed out by Koba and Takeda that their counter term $(e^2/3\pi^2)\mathbf{A}^T(x)^2 \int^\infty k dk$ is not invariant under a Lorentz transformation. The above analysis shows how its corresponding term in (92) can be derived from a covariant theory when this theory involves divergence and ambiguity.

From equations (87) and (76) it can be seen that the inclusion of the counter term $(e^2/3\pi^2)\mathbf{A}^2 \int^\infty k dk$ in the Hamiltonian may be interpreted as the subtraction of the photon self energy. (Cf. Corinaldesi and Jost 1948).

Calculations similar to that in the present section have recently been made by Corinaldesi and Field (1949) for the self energy of the meson due to interaction with nucleons and by Gupta (1949) for the self energy of the electron due to interaction with the electromagnetic field.

Note added in proof.—Evaluation of the induction tensor by means of the Fourier integral representation of the Δ -functions has also been given by Pauli and Villars in connection with the problem of regularization and more recently by Schwinger (1949, Part III., Appendix). The methods of these authors differ from the method of this section in that the tensor form is used from the beginning. The result of Pauli and Villars, if it is not regularized, involves the same two divergent integrals I and J as in the present work, and also a singularity on the cone $p_\lambda^2=0$ which gives rise to a photon self energy in agreement with our result (87). Schwinger on the other hand, has re-established his conclusions about the gauge invariance and the vanishing of the photon self energy. In view of the ambiguity in handling divergent integrals, it is not to be expected that the results obtained by the different methods should agree with one another.

§ 4. THE ENERGY-MOMENTUM REPRESENTATION.

Since an arbitrary electromagnetic field can be resolved by Fourier analysis into periodic components, we can confine our attention to a potential of the form

$$A_\mu(x) = a_\mu \exp(ip_\lambda x_\lambda). \quad (94)$$

In this case the general formula (18) reduces to

$$\delta j_\mu(x) = -4e^2 K_{\mu\nu}(p) A_\nu(x). \quad (95)$$

It is instructive to derive this formula from the ordinary form of perturbation theory developed for the energy momentum representation of quantum electrodynamics. We consider discrete states of the electron and electromagnetic waves enclosed in a volume V . Let a_r, b_r be the constant spinors for the states of the electron waves with momentum \mathbf{k}_r and energies $\pm E_r$, where $E_r = (\mathbf{k}_r^2 + m^2)^{\frac{1}{2}}$. Let α and β be the Dirac matrices and $\alpha_\mu = (\alpha, i)$.

Consider first the induced current due to the polarization of the vacuum by a periodic external field $A_\mu^{\text{ext}}(x)$. For simplicity we consider here the case when A_μ^{ext} is constant with respect to the time. The interaction between $A_\mu^{\text{ext}}(x)$ and the electric current $j_\mu(x)$ is given by

$$H^{\text{ext}} = - \int j_\mu(x) A_\mu^{\text{ext}}(x) d^3x. \quad (96)$$

The expectation value of the induced current is

$$\delta j_\mu(x) = - \sum_{r,s} \frac{1}{E_r + E_s} \{ (o | H^{\text{ext}} | rs) (rs | j_\mu(x) | o) + (o | j_\mu(x) | rs) (rs | H^{\text{ext}} | o) \}, \quad (97)$$

where r, s denote the states of the virtual electron pairs. It follows from (96), (97), that

$$\delta j_\mu(x) = \frac{e^2}{V} A_\nu^{\text{ext}}(x) \sum_{r,s} \frac{1}{E_r + E_s} (b_r^\dagger \alpha_\nu a_s a_s^\dagger \alpha_\mu b_r + b_s^\dagger \alpha_\mu a_r a_r^\dagger \alpha_\nu b_s) \delta(\mathbf{p}, \mathbf{k}_r - \mathbf{k}_s), \quad (98)$$

where we use the notation

$$\begin{aligned}\delta(\mathbf{p}, \mathbf{q}) &= 1 \quad \text{for } \mathbf{p} = \mathbf{q} \\ &= 0 \quad \text{for } \mathbf{p} \neq \mathbf{q}.\end{aligned}$$

Summing over the spins, we obtain

$$\delta j_\mu(x) = \frac{e^2}{V} A_\nu^{\text{ext}}(x) \sum_{r,s} \frac{1}{E_r + E_s} \text{Tr}(\alpha_\nu A_s^+ \alpha_\mu A_r^- + \alpha_\mu A_r^+ \alpha_\nu A_s^-) \delta(\mathbf{p}, \mathbf{k}_r - \mathbf{k}_s) \quad (99)$$

with

$$A_r^\pm = \frac{1}{2} [1 \pm (\boldsymbol{\alpha} \cdot \mathbf{k}_r + \beta m) / E_r]. \quad (100)$$

Evaluation of the traces gives

$$\begin{aligned}\delta j_\mu(x) &= \frac{e^2}{V} A_\nu^{\text{ext}}(x) \sum_{r,s} \frac{1}{E_r E_s (E_r + E_s)} [r_\mu s_\nu + r_\nu s_\mu + \delta_{\mu\nu} (-r_\lambda s_\lambda + m^2)] \\ &\quad \times [\delta(\mathbf{p}, -\mathbf{k}_r - \mathbf{k}_s) + \delta(\mathbf{p}, \mathbf{k}_r + \mathbf{k}_s)], \quad (101)\end{aligned}$$

with the notation $r_\mu = (\mathbf{k}_r, iE_r)$. We replace here the summation by an integration as usual. This gives

$$\begin{aligned}\delta j_\mu(x) &= \frac{e^2}{(2\pi)^3} A_\nu^{\text{ext}}(x) \iint \frac{1}{EE'(E+E')} [q_\mu q'_\nu + q_\nu q'_\mu + \delta_{\mu\nu} (-q_\lambda q'_\lambda + m^2)] \\ &\quad \times [\delta(\mathbf{p} + \mathbf{q} + \mathbf{q}') + \delta(\mathbf{p} - \mathbf{q} - \mathbf{q}')] d^3q d^3q', \quad (102)\end{aligned}$$

which is just equation (95) with $K_{\mu\nu}(p)$ given by (43) and $p_\lambda = (\mathbf{p}, 0)$.

We consider next the induced current due to the interaction of a radiating electron with the vacuum electrons through their interaction with the electromagnetic field. This interaction is given by the Hamiltonian

$$H = H^{(1)} + H^{(2)}, \quad (103)$$

where $H^{(1)}$ gives the interaction of the electron waves with the transverse electromagnetic field and $H^{(2)}$ gives the Coulomb interaction of the electron waves that arises from their interaction with the longitudinal electromagnetic field. (See Wentzel 1943.)

The induced current is given by

$$\delta j_\mu(x) = S^{(2)\dagger} j_\mu(x) + S^{(1)\dagger} j_\mu(x) S^{(1)} + j_\mu(x) S^{(2)}, \quad (104)$$

where, as shown by Heitler and the present writer (1949),

$$S_{mn}^{(1)} = \frac{H_{mn}^{(1)}}{E_n - E_m}, \quad S_{mn}^{(2)} = \frac{1}{E_n - E_m} \left(H_{mn}^{(2)} + \sum_k \frac{H_{mk}^{(1)} H_{kn}^{(1)}}{E_n - E_k} \right). \quad (105)$$

We consider the matrix element of $\delta j_\mu(x)$ for a transition of an electron from a state A to a state B of the same energy, $E_A = E_B$. Let $\mathbf{k}_A, \mathbf{k}_B$ be the initial and final values of the momentum of the electron, $\Delta \mathbf{k} = \mathbf{k}_A - \mathbf{k}_B$. Taking account of all the intermediate states of the vacuum polarization

type (See Ito, Koba and Tomonaga, 1948), we obtain

$$[\delta j_\mu(x)]_{BA} = \frac{4\pi e^3}{V^2 \Delta \mathbf{k}^2} \exp(i \Delta \mathbf{k} \cdot \mathbf{x}) \sum_{r,s,T} \delta(\mathbf{k}_r - \mathbf{k}_s, \Delta \mathbf{k}) \frac{1}{E_r + E_s} \{a_s^\dagger \alpha_\mu b_r (b_r^\dagger a_s a_A^\dagger a_A - b_r^\dagger a_T a_s a_B^\dagger \alpha_T a_A) + b_s^\dagger \alpha_\mu a_r (a_r^\dagger b_s a_B^\dagger a_A - a_r^\dagger \alpha_T b_s a_B^\dagger \alpha_T a_A)\}, \quad (106)$$

where the index T denotes the direction of polarization of a photon in an intermediate state. Proceeding in the same way as in the derivation of the Møller interaction (See Heitler 1944) we can reduce (106) to the form

$$[\delta j_\mu(x)]_{BA} = \frac{-4\pi e^3}{V^2 \Delta \mathbf{k}^2} \exp(i \Delta \mathbf{k} \cdot \mathbf{x}) \sum_{r,s} \delta(\mathbf{k}_r - \mathbf{k}_s, \Delta \mathbf{k}) \frac{1}{E_r + E_s} (a_s^\dagger \alpha_\mu b_r b_r^\dagger \alpha_\nu a_s + b_s^\dagger \alpha_\mu a_r a_r^\dagger \alpha_\nu b_s) a_B^\dagger \alpha_\nu a_A, \quad (107)$$

where ν runs from 1 to 4. Now

$$[j_\mu(x)]_{BA} = \frac{-e}{V} a_B^\dagger \alpha_\mu a_A \exp(i \Delta \mathbf{k} \cdot \mathbf{x}), \quad (108)$$

so that the matrix elements of the electromagnetic potentials describing the field of the electron are, by (20),

$$[A_\mu(x)]_{BA} = \frac{-4\pi e}{V \Delta \mathbf{k}^2} a_B^\dagger \alpha_\mu a_A \exp(i \Delta \mathbf{k} \cdot \mathbf{x}). \quad (109)$$

Hence

$$[\delta j_\mu(x)]_{BA} = \frac{e^2}{V} [A_\nu(x)]_{BA} \sum_{r,s} \delta(\mathbf{k}_r - \mathbf{k}_s, \Delta \mathbf{k}) \frac{1}{E_r + E_s} (a_s^\dagger \alpha_\mu b_r b_r^\dagger \alpha_\nu a_s + b_s^\dagger \alpha_\mu a_r a_r^\dagger \alpha_\nu b_s). \quad (110)$$

This result is of the same form as (98) and can be dealt with in the same way. It leads to the result (95) with $K_{\mu\nu}(p)$ given by (43) and $p_\lambda = (\Delta \mathbf{k}, 0)$.

Finally we observe that the energy level shift of the vacuum and the self energy of the photon due to the creation of virtual electron pairs can be calculated by means of the general formula

$$\Delta E_n = \sum_m \frac{H_{nm}^{(1)} H_{mn}^{(1)}}{E_n - E_m}, \quad (111)$$

and the formulas in this section. The result for the photon self energy thus obtained is given by formula (24) with $K_{\mu\nu}(p)$ given by (43).

§ 5. REGULARIZATION.

Pauli and Villars have applied the regularization to several forms of $K_{\mu\nu}(p)$. We shall now apply their method to the expressions for $K_{\mu\nu}(p)$ given by (44).

Applied to this problem, the rules of regularization may be stated as follows:

(i) We first perform the integration in (44) over a finite region in the \mathbf{k} -space. It will be convenient for us to take the domain of integration to be a sphere of radius \bar{k} . We denote the dependence of $K_{\mu\nu}(p)$ on \bar{k} and the mass m by writing it in the form $K_{\mu\nu}(p, \bar{k}, m^2)$.

(ii) The above $K_{\mu\nu}(p)$ is replaced by

$$\widetilde{K}_{\mu\nu}(p)=\sum_i C_i K_{\mu\nu}(p, k, M_i^2) \quad . \quad . \quad . \quad . \quad . \quad (112)$$

with $i=0, 1, 2, \dots$ and $C_0=1, M_0=m$. The coefficients C_i should satisfy the conditions

$$\sum_i C_i=0, \quad . \quad . \quad . \quad . \quad . \quad (113)$$

$$\sum_i C_i M_i^2=0, \quad . \quad . \quad . \quad . \quad . \quad (114)$$

$$\sum_i C_i \log M_i^2=0. \quad . \quad . \quad . \quad . \quad . \quad (115)$$

The first two conditions are necessary for ensuring the gauge invariance. The third condition is necessary for eliminating the self charge.

(iii) \bar{k} is now increased to infinity.

(iv) The auxiliary masses M_i ($i=1, 2, \dots$) are subsequently increased to infinity in such a way that the conditions (113), (114), (115) are preserved and

$$\sum_{i \neq 0} (C_i/M_i^2) \rightarrow 0. \quad . \quad . \quad . \quad . \quad . \quad (116)$$

Our first step is to perform the integration in (44) over the sphere $|\mathbf{k}| \leq \bar{k}$. This sphere can be divided into two domains according to whether the surface $w=\text{constant}$ is completely or only partially enclosed in the sphere. The corresponding domains in the $\phi y z$ -space are

$$(I) \quad 0 \leq \phi \leq 2\pi, \quad 0 \leq z \leq z_a, \quad -z \leq y \leq z, \quad . \quad . \quad . \quad (117)$$

$$(II) \quad 0 \leq \phi \leq 2\pi, \quad z_a \leq z \leq z_b, \quad -u(z) \leq y \leq u(z). \quad . \quad (118)$$

Here z_a and z_b are defined by the equations

$$\bar{k}^2 = \frac{1}{4} \mathbf{p}^2 z_a^2 + \frac{m^2 z_a^2}{1-z_a^2}, \quad . \quad . \quad . \quad . \quad . \quad (119)$$

$$\bar{k}^2 = \frac{m^2 z_b^2}{1-z_b^2}, \quad . \quad . \quad . \quad . \quad . \quad (120)$$

and $u(z)$ is defined by

$$\frac{1}{4} \mathbf{p}^2 u(z)^2 = \bar{k}^2 - \frac{m^2 z^2}{1-z^2}. \quad . \quad . \quad . \quad . \quad . \quad (121)$$

As \bar{k} increases to infinity, z_a and z_b both tend to unity and the limits of integration tend to the values given by (53). It follows from the above equations that, for any function $\psi(z)$ of z ,

$$\int_{|\mathbf{k}| \leq \bar{k}} \psi(z) d^3 k = 4\pi \int_0^{z_a} z \psi(z) dz + \frac{\pi \mathbf{p}^2}{m^2} \int_0^{z_a} u^2 \frac{\psi(z)(1-z^2)^2}{z} du, \quad . \quad . \quad (122)$$

$$\begin{aligned} \int_{|\mathbf{k}| \leq \bar{k}} \psi(z)(1-y^2) d^3 k &= 4\pi \int_0^{z_a} (z - \frac{1}{3} z^3) \psi(z) dz \\ &+ \frac{\pi \mathbf{p}^2}{m^2} \int_0^{z_a} (u^2 - \frac{1}{3} u^4) \frac{\psi(z)(1-z^2)^2}{z} du. \quad . \quad (123) \end{aligned}$$

In the integrals with respect to u the factor involving the variable z should be expressed in terms of the variable u by means of (121).

As shown in § 3, the triple integrals $K_{\mu\nu}(p)$ can be reduced to the double integrals R, S, T. Integrating over the domains (117), (118) and using the formulas (122), (123), we obtain

$$R=3(4\pi)^{-2}[p_\lambda^2(J_a+f_a)-4I_a]-\mathbf{p}^2/24\pi^2+0(1/\bar{k}), \quad . \quad . \quad (124)$$

$$S=-(2\pi)^{-2}(J_a+f_a)+0(1/\bar{k}), \quad . \quad . \quad . \quad . \quad . \quad (125)$$

$$T=(4\pi)^{-2}[p_\lambda^2(J_a+f_a)-4I_a]-\mathbf{p}^2/60\pi^2+0(1/\bar{k}). \quad . \quad . \quad . \quad (126)$$

Here I_a , J_a , f_a are the integrals given by (66), (67), (27) with the upper limit z_a instead of unity, and $0(1/\bar{k})$ denotes terms which vanish when $\bar{k} \rightarrow \infty$. It follows from (66), (67), that

$$I_a = \frac{m^2}{3} \frac{z_a^3}{1-z_a^2}, \quad . \quad . \quad . \quad . \quad . \quad (127)$$

$$J_a = \frac{1}{3} \left(\log \frac{1+z_a}{1-z_a} - 2z_a + \frac{1}{3}z_a^3 \right), \quad . \quad . \quad . \quad . \quad . \quad (128)$$

so that, by (119)

$$I_a = \frac{1}{3}(\bar{k}^2 - \frac{1}{4}\mathbf{p}^2 - \frac{1}{2}m^2) + 0(1/\bar{k}), \quad . \quad . \quad . \quad . \quad . \quad (129)$$

$$J_a = \frac{1}{3}(\log 4 - \frac{5}{3} - \log m^2 + \log \bar{k}^2) + 0(1/\bar{k}). \quad . \quad . \quad . \quad . \quad (130)$$

Hence

$$\widetilde{I} = \frac{1}{3}(\bar{k}^2 - \frac{1}{4}\mathbf{p}^2) \Sigma C_i - \frac{1}{6} \Sigma C_i M_i^2 + 0(1/\bar{k}), \quad . \quad . \quad . \quad . \quad (131)$$

$$\widetilde{J} = \frac{1}{3}(\log 4 - \frac{5}{3} + \log \bar{k}^2) \Sigma C_i - \frac{1}{3} \Sigma C_i \log M_i^2 + 0(1/\bar{k}). \quad . \quad . \quad (132)$$

On account of the conditions (113), (114), (115), we have

$$\widetilde{I} = 0(1/\bar{k}), \quad \widetilde{J} = 0(1/\bar{k}), \quad . \quad . \quad . \quad . \quad (133)$$

and hence

$$\widetilde{R} = 3(4\pi)^{-2} p_\lambda^2 \widetilde{f} + 0(1/\bar{k}), \quad . \quad . \quad . \quad . \quad (134)$$

$$\widetilde{S} = -(2\pi)^{-2} \widetilde{f} + 0(1/\bar{k}), \quad . \quad . \quad . \quad . \quad (135)$$

$$\widetilde{T} = (4\pi)^{-2} p_\lambda^2 \widetilde{f} + 0(1/\bar{k}). \quad . \quad . \quad . \quad . \quad (136)$$

We now make \bar{k} increase to infinity and subsequently make the auxiliary masses increase to infinity according to the rules. This gives

$$\widetilde{I} = 0, \quad \widetilde{J} = 0, \quad . \quad . \quad . \quad . \quad (137)$$

$$\widetilde{R} = 3(4\pi)^{-2} p_\lambda^2 f, \quad \widetilde{S} = -(2\pi)^{-2} f, \quad \widetilde{T} = (4\pi)^{-2} p_\lambda^2 f, \quad . \quad (138)$$

where f is the finite and invariant function of p_λ given by (27). Substituting (138) into (58), (59), we obtain

$$\widetilde{K}_{\mu\nu}(p) = (4\pi)^{-2} (\delta_{\mu\nu} p_\lambda^2 - p_\mu p_\nu) f(p), \quad . \quad . \quad . \quad (139)$$

CIII. *Study of the Wave-forms of Atmospherics.*

By S. R. KHASTGIR, D.Sc.(Edin.), F.N.I.*, and R. ROY, M.Sc.†‡.

[Received June 13, 1949.]

§ 1. INTRODUCTION.

SOME experiments to study the wave-forms of distant atmospherics were carried out in the Physics Laboratory of the Dacca University in the monsoon months of 1947. The method consisted in examining by means of a cathode ray oscillograph worked with linear time base the potential variations developed across a condenser in a damped antenna, the time-constant of which was kept small compared with the duration of the atmospheric pulse. All the wave-forms observed in this manner have been broadly classified. In addition to what were already observed by previous workers on the subject, viz.,

- (1) Aperiodic and *quasi*-periodic pulses,
- (2) Aperiodic or *quasi*-periodic pulse followed by a slow component,
- (3) "Precursors" of relatively higher *quasi*-frequencies,
- (4) "Repeater" pulses and wavy border on the slow components due to multiple reflection between the earth and the ionosphere,

the following observations were made :—

- (i) Slow periodic components of an average duration of about 3 milli-seconds per cycle,
- (ii) Aperiodic or *quasi*-periodic pulse preceded by a slow component,
- (iii) Aperiodic or *quasi*-periodic pulse preceded and followed by slow components.

The results of the experimental investigation with an account of the experimental arrangement and procedure are given in this paper. Interpretations of the observed wave-forms have also been suggested.

§ 2. EQUIPMENT AND EXPERIMENTAL ARRANGEMENT.

1. *Aerial Unit.*

One open-air horizontal aerial 70 metres long with an effective height of 16.25 metres from the ground was installed for the study of wave-forms.

A resistance R ($0.05\text{ M}\Omega$) and a condenser C_1 ($0.01\text{ }\mu\text{F}$) were placed in series with the aerial to render the latter highly aperiodic. The output across the condenser C_1 for the study of the variation of E and that across the resistance R for the study of variations of dE/dt could be applied to the amplifier input by means of a change-over arrangement. The resistance R and the condenser C_1 were enclosed in a well-shielded box.

* Benares Hindu University,

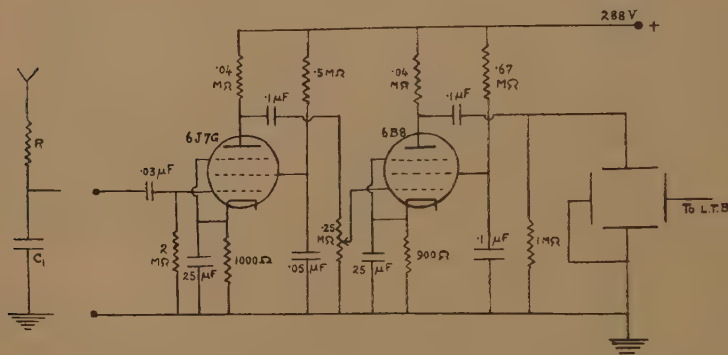
† College of Science, Calcutta University.

‡ Communicated by the Authors.

2. Amplifier Unit.

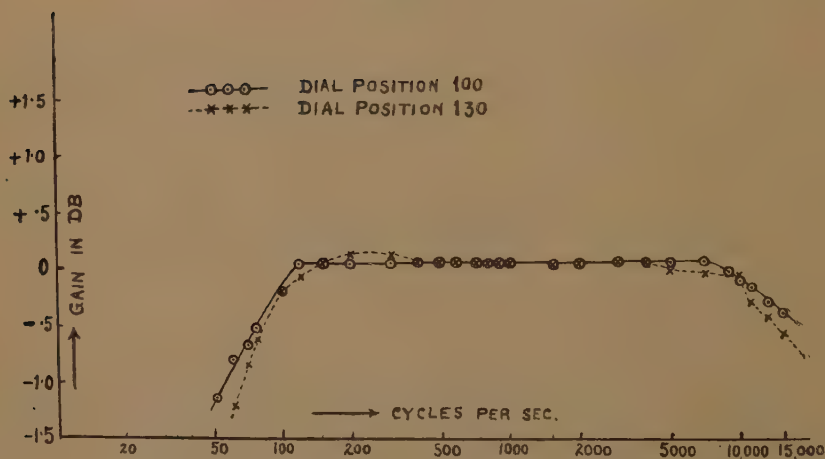
One Class A-operated audio-amplifier having a flat gain characteristic with minimum distortion for a wide range of frequencies was constructed. As the atmospherics do not occur below 100 cycles, the resistance-capacity-coupled audio-amplifier appeared to be the most advantageous and suitable for the purpose. The circuit diagram of the R-C amplifier is shown in fig. 1.

Fig. 1.



Circuit diagram of the audio amplifier.

Fig. 2.



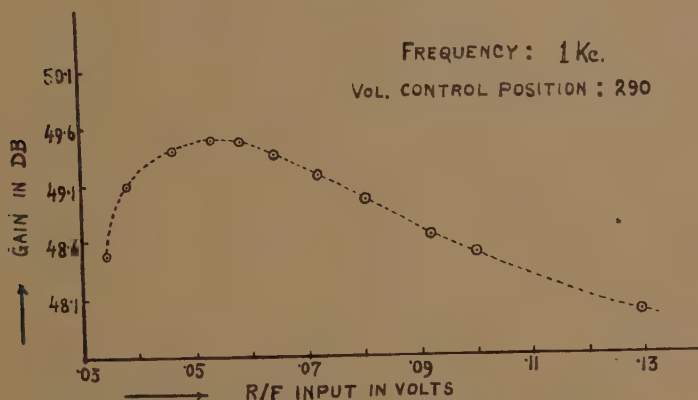
Frequency response curves for the amplifier.

For manual control of the total gain of the amplifier the grid leak of the second stage consisted of a potentiometer having a graduated dial with an indicator. The frequency-response curves of the amplifier showing voltage-gain relative to the gain at 1 kc. in db for different frequencies are given in fig. 2, for two different gain-levels, *i. e.* for two different positions.

of the volume control. The voltage-gain was within ± 0.5 db from 90 cycles to 12 kc./s. A typical curve showing the voltage-gain for different input voltages for a fixed gain level at 1 kc. frequency is shown in fig. 3.

The time-constant of the input circuit of the amplifier unit was found to be 21,800 microseconds. Since the time-constant was large in comparison with the average duration of the pulse, no sensible distortion was possible.

Fig. 3.



Voltage-gain for different input voltages at 1 Kc.

3. Oscillograph Unit and Linear Time Base.

A Cossor 09D model double beam oscillograph was employed. The output of the amplifier provided the Y—or vertical deflection—and for the delineation of the atmospheric pulses, the linear time base was used along the X—or horizontal direction. The oscillograph screen was provided with graduations in cm. and mm. marked radially from the undeflected position of the oscillograph spot. The deflection-sensitivity of the oscillograph was found by noting the sweep-length of the spot corresponding to the 50-cycle known voltages obtained from a potentiometric device. The sweep-length was found to increase linearly with the applied voltage over the experimental range.

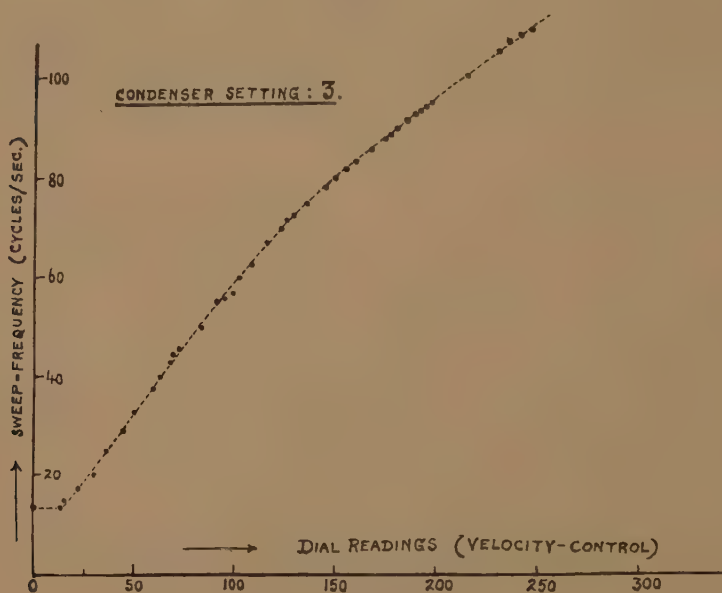
4. Time Base Calibration of the Oscillograph.

The standardized audio tone of 1 kc./s. of the local transmitting station was used to calibrate the oscillograph time base by counting the number of waves on the screen, when the pattern appeared stationary. Additional calibration points were obtained with the aid of a beat frequency oscillator, the frequency of which could be adjusted to rational fractions of 1000 cycles, also by the stationary pattern method.

During calibration or during the course of observations, the amplitude-control was set for the maximum sweep-length, and the trigger for minimum duration of fly-back stroke.

A typical calibration graph showing sweep-frequency for the different positions of the velocity-control corresponding to one condenser setting is shown in fig. 4.

Fig. 4.



Time-base calibration.

5. Measurement of the Gain-Characteristic of the Amplifier Unit.

The gain characteristic of the aperiodic amplifier was determined in the usual way. The output of the beat-frequency oscillator set at some frequency was connected to a step-down transformer and the voltage across a potentiometer in the secondary circuit was applied to the input terminals of the amplifier unit, the output terminals of the latter having been connected to the pair of deflecting plates of the oscillograph. For a given input voltage, as measured by a previously calibrated valve-voltmeter, the sweep-length on the oscillograph for the particular frequency was noted. From the deflection-sensitivity curve of the oscillograph, the output voltage of the amplifier corresponding to the observed sweep-length was found. The output voltages corresponding to the same input voltage at different frequencies were in this way found and the ratio of the output voltage at any frequency to that at 1 kc. expressed in db was determined for different frequencies.

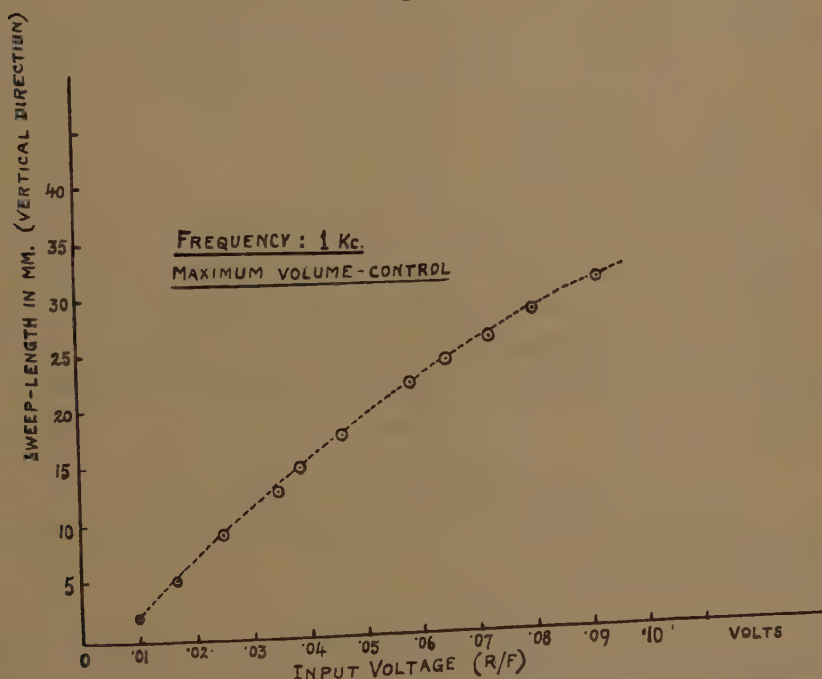
§3. EXPERIMENTAL PROCEDURE.

The variations of the electric field due to distant atmospheric pulses were studied by observing the voltage variations across the condenser C_1 placed in series with the aerial. The potential variations appeared as wave-forms on the oscillographic screen. The wave-forms were carefully traced on suitable squared paper. Observations were usually made

during the night hours between 11 P.M. to 1 A.M. I.S.T. Before starting to trace the wave-forms on any particular run, the time-base condenser and the velocity-control were adjusted to suit the duration of the transients which appeared to be predominant at the time of the observations. The wave-forms showing the E-variations only were studied. The variations of dE/dt as obtained by applying the voltage across the resistance in the aerial to the input of the amplifier (and subsequently to the deflecting plates of the cathode ray oscillograph after suitable amplification) were far too complex to be delineated by the "eye-and-hand" method.

For the determination of the actual values of field strength of the atmospheric pulses, a calibration curve showing the observed vertical sweep-lengths on the oscillograph for different input voltages to the amplifier was drawn (see fig. 5). The data necessary for this curve were

Fig. 5.



obtained during the measurement of the voltage-gain of the amplifier for different input voltages. Having obtained from the calibration curve the value of the amplifier input voltage V corresponding to any vertical displacement of the wave-form tracing, the field strength was found from $E = V/h_e$, where h_e is the effective height of the aerial.

Since the audio-amplifier was a two-stage unit it is to be noted that the output voltage of the amplifier was in phase with the amplifier input voltage. Thus the observed *downward* deflection of the oscillograph spot when a positive potential was applied to the vertical plate, indicated a positive value of the charge on the upper plate of the aerial condenser and a positive change of the electric field.

§4. EXPERIMENTAL RESULTS.

433 oscillograms of distant atmospherics were recorded. They are classified as follows :

- Type I. Slow aperiodic pulses with or without change to the opposite cycle. Number observed : 63.
- Type II. *Quasi*-periodic pulses (damped oscillation type). Number : 112.
- Type III. Aperiodic or *quasi*-periodic pulse followed by a slow component. Number : 171.
- Type IV. Slow component preceding an aperiodic or *quasi*-periodic pulse. Number : 16.
- Type V. Slow components on both sides of an aperiodic or *quasi*-periodic pulse. Number : 9.
- Type VI. Slow periodic components. Number : 48.
- Type VII. Oscillograms showing multiple reflection between the earth and the ionosphere in the form of "repeater" pulses or wavy borders. Number of "repeater" pulses : 4.
- Type VIII. Mixed variety. Number : 10.

Typical oscillograms of each type are shown in figs. 6 and 7. A number of wave-forms showed high frequency embroideries. The types IV., V. and VI. were not observed by other investigators.

The details of the different types, viz., total number of individual types, average duration, superposed ripple frequency, percentage of positive and negative aperiodic and *quasi*-periodic pulses, average field-strengths etc., are given in Tables I.-VI.

TABLE I.

Type I.—Aperiodic pulses.

Total number=63. Percentage of occurrence=11.5.
 Percentage of *positive* aperiodic—82.5.
 Percentage of *negative* aperiodic—17.5.
 Duration : Average—1.71 msec. ; Maximum—6.95 msec. ; Minimum—300 μ sec.
 Superposed ripple frequency—3.7 kc./s. to 7.8 kc./s.
 Field-Strength : Average—20.6 mv./metre ; Maximum—56.6 mv./metre ;
 Minimum—7.6 mv./metre.

TABLE II.

Type II.—*Quasi*-periodic pulses.

Total number=112. Percentage of occurrence=25.9.
 Maximum number of cycles=8.
 Duration : Average—108 msec. ; Maximum—2.05 msec. ; Minimum—98.5 μ sec.
 Average frequency of each cycle=2.646 kc./s.
 Field-strength : Average=24.9 mv./metre ; Maximum=51.6 mv./metre ;
 Minimum=7.6 mv./metre.

Fig. 6.

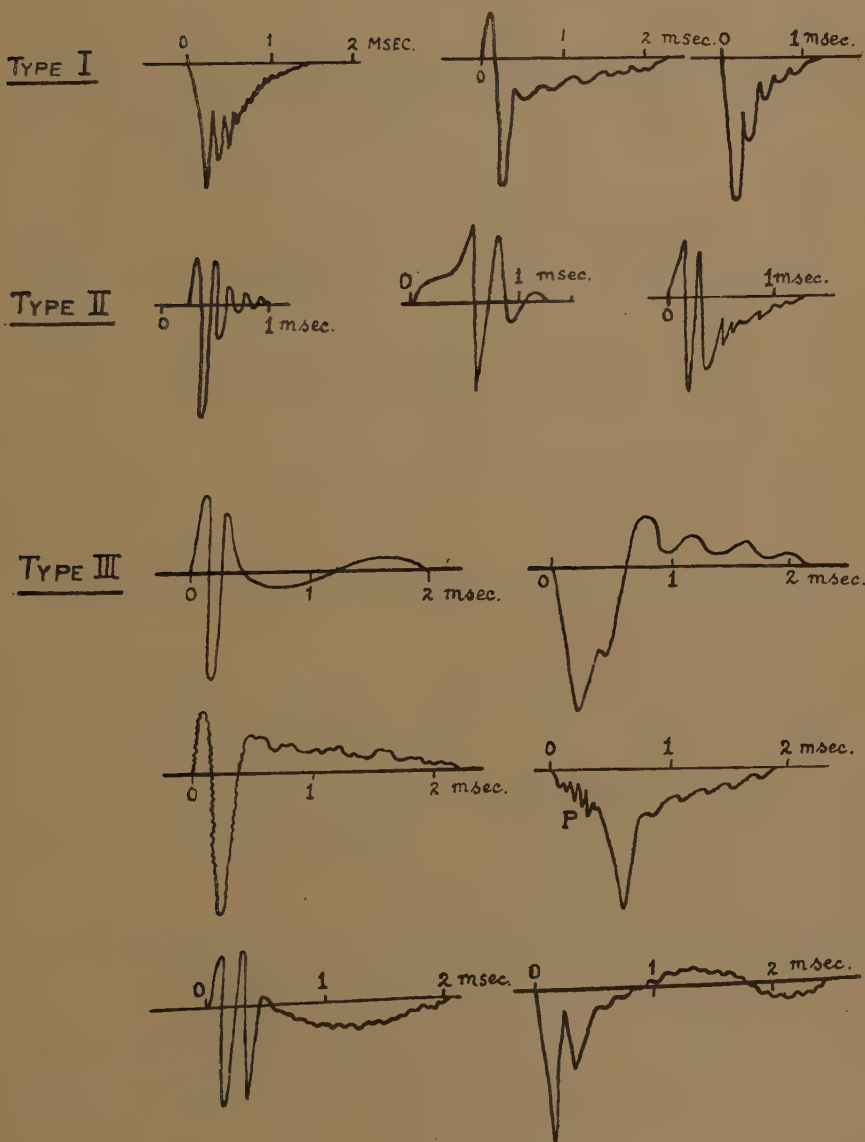
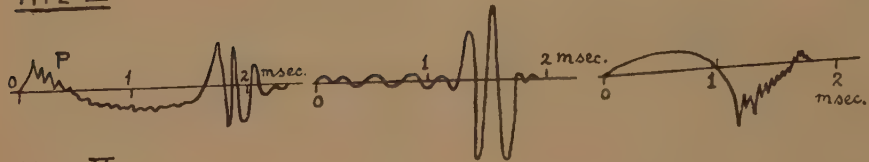
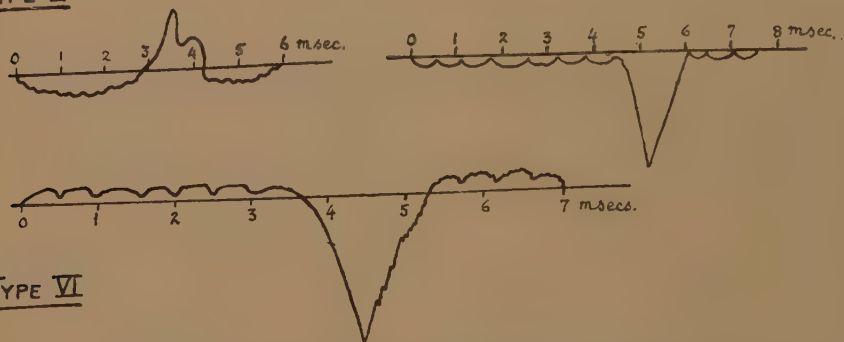


Fig. 7:

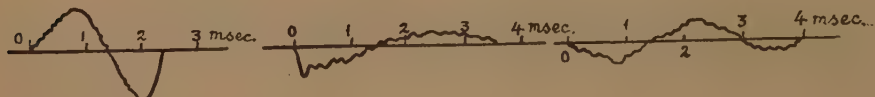
TYPE IV



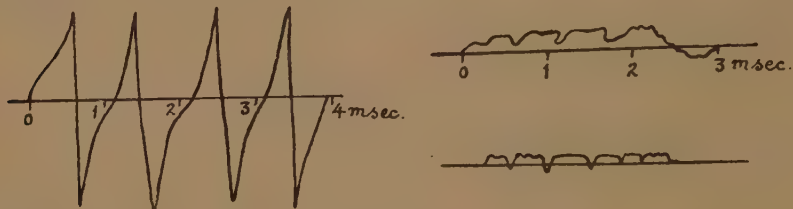
TYPE V



TYPE VI



TYPE VII



TYPE VIII

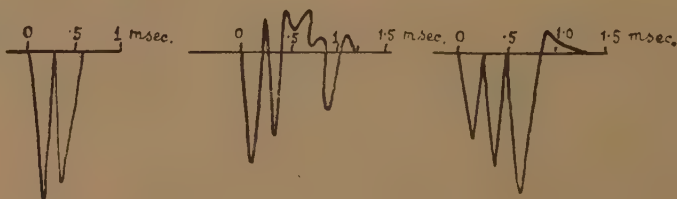


TABLE III A.

Type III A.—*Quasi-periodic* pulse followed by a slow component.

Total number=81. Percentage of occurrence=18.7.

Maximum number of associated cycles=3.

Average total duration=2.03 msec.

Duration of Slow component: Average=1.48 msec.; Maximum=4.55 msec.;
Minimum=300 μ sec.

Duration of *quasi-periodic*: Maximum=2.28 msec.; Minimum=158 μ sec.

Average frequency of *quasi-periodic*=1.66 kc./s.

Average maximum field-strength of *quasi-periodic*=27.9 mv./metre.

Average maximum field-strength of slow component=10.5 mv./metre.

Field-strength of slow component: Maximum=22.9 mv./metre; Minimum
=6.4 mv./metre.

Percentage of slow component of +ve type=51.8.

Percentage of slow component of -ve type=6.2.

Percentage of slow component of \pm or \mp type=42.

Frequency of ripples superposed on slow components=3.75 kc./s. to 4 kc./s.

Number of times "precursors" observed in this type=2.

TABLE III B.

Type III B.—*Aperiodic* pulse followed by a slow component.

Total number=90. Percentage of occurrence=20.7.

Average total duration=2.51 msec.

Duration of slow component: Average=1.78 msec.; Maximum=4.86 msec.;
Minimum=276.7 μ sec.

Duration of aperiodic: Average=732.4 μ sec.; Maximum=2.17 msec.;
Minimum=79.1 μ sec.

Field-strength of aperiodic: Average maximum=32.2 mv./metre; Maximum
=56.6 mv./metre; Minimum=9.0 mv./metre.

Field-strength of slow component: Average maximum=12.2 mv./metre;
Maximum=30.3 mv./metre; Minimum=6.3 mv./metre.

Percentage of + aperiodic followed by + slow component=37.8; of + aperiodic
followed by - slow component=53.4; of + aperiodic followed by
 \pm slow component=7.8; of - aperiodic followed by + slow com-
ponent=1.0.

Total duration of "precursor"=291.8 μ sec.

Frequency of "precursor"=6.86 kc./s.

Number of cycles in "precursor"=2.

Number of times "precursors" observed in this type=2.

TABLE IV.

Type IV.—Slow component preceding an aperiodic or *quasi*-periodic pulse.

Total number=16.

Total duration: Average=3.69 msec.; Maximum=9.26 msec.; Minimum=650 μ sec.

Duration of slow component: Average=2.17 msec.; Maximum=5.09 msec.; Minimum=300 μ sec.

Percentage of + slow component associated with + A=12.4; of + slow component associated with - A=6.2; of - slow component associated with + A=37.6; of \pm slow component associated with + A=6.2; of - slow component associated with Q=37.6; of + slow component associated with Q=0.0.

Average field-strength associated with Q or A=9.1 mv./metre.

Average field-strength of slow component=28.5 mv./metre.

Total duration of "precursor"=322.5 μ sec.

Duration between "precursor" and *quasi*-periodic=5.19 msec.

Frequency of "precursor"=12.9 kc./s.

Number of cycles in "precursor"=3.

Number of times "precursors" observed in this type=1.

TABLE V.

Type V.—Slow components on both sides of an aperiodic or *quasi*-periodic pulse.

Total number=9. Average total duration=2.67 msec.

Duration of preceding slow component: Mean=1.27 msec.; Maximum=3.24 msec.

Duration of succeeding slow component: Mean=742.9 μ sec.; Maximum=1.59 msec.

Percentage of (- slow)+(+ aperiodic)+(- slow)=55.6; of (+ slow)+(+ aperiodic)+(+ slow)=33.3; of (- slow)+(*quasi*-periodic)+(+ slow)=11.1.

Average maximum field-strength of preceding slow component=8.6 mv./metre.

Average maximum field-strength of A or Q=36.2 mv./metre.

Average maximum field-strength of succeeding slow component=6.6 mv./metre.

TABLE VI.

Type VI.—Slow periodic pulses.

Total number=48. Percentage of occurrence=11.1.

Average duration =3.01 msec. per cycle.

Frequency range=130.4 to 1265 cycles.

Superposed ripple frequency=10 kc./s. to 3.33 kc./s.

Field-strength: Average=10.46 mv./metre; Maximum=30.29 mv./metre; Minimum=7.52 mv./metre.

The average durations of the slow components associated with the aperiodic or *quasi*-periodic pulses, of the slow periodic pulses per cycle, of the observed aperiodic and *quasi*-periodic pulses, "precursors" and superposed ripples are given in Table VII. for comparison.

It is to be observed that the aperiodic types, and also the associated slow components have been classified according to sign. Following the convention adopted by Appleton, the *positive* sign is used to indicate that the predominant change of field was in such a direction as to give a positive charge to the upper plate of the condenser C_1 , and that the field tended to send positive electricity down the aerial to earth. Thus in the A^+ type the upper plate of the condenser became more positive and then returned to its initial condition. Considering the aperiodic pulses it was found.

TABLE VII.

Impulses	Average duration in msec.	Average frequency
Slow periodic pulses (Type VI.)	3.01	323 cycles/sec.
Slow components	1.51	—
Aperiodics	1.71	—
<i>Quasi</i> -periodics	0.38	2.65 kc./s.
Precursors	0.11	8.95 kc./s.
Ripples	0.20	5 kc./s.

that the net changes of the earth's field were more frequently positive than negative in sign. In the cases of the slow components following or/and preceding the aperiodic or *quasi*-periodic pulses, the net changes of the field were, however, found more negative.

§ 5. GENERAL CONCLUSIONS AND INTERPRETATIONS OF THE WAVE-FORMS.

It had been shown by Walter and others (1903, 1935) from moving camera observations that a single lightning flash to earth was made up of a downward *leader* stroke followed by an upward faster *return* stroke, the duration of the processes being of the order of 10 and 0.05 msec. respectively. The work of Schonland, Malan and Collens (1935) on lightning discharges confirmed these conclusions. Appleton and Chapman (1937) were able to identify the leader stroke as causing the initial slow change of field (the "*a*"-stage) and the return stroke as producing the most rapid change of field (the "*b*"-stage) after the initial slow change. It is now accepted that the leader streamer of the first stroke lowers negative charge into the air and distributes it over the conducting system formed by the leader channel causing a considerable change of thundercloud moment. In the return stage of the discharge there is an upward passage of positive streamers from the earth just before the arrival of one of the leader branches at the ground. These streamers were photographed by McEachron and McMorris (1937) and the frequent presence of more than one was evident from the root branching shown at the base of many lightning channels photographed by Schonland, Malan and Collens (1935).

As was first shown by Simpson (1926), the positive return streamer rapidly advances by drawing electrons, produced by collision in front of its tip and passing them down to ground *via* its conducting stem, and there is a large and rapid change of electric moment which is associated with the "b"-stage of Appleton and Chapman. The last portion of the return stroke involves the removal and passage down the channel of the residual charge on the cloud centre tapped by the stroke. This stage is marked by a continuance of channel luminosity after the return streamer has entered the cloud and is of comparatively long duration. According to Schonland (1938), this may be identified with the final slow field change (the "c"-stage) observed by Appleton and Chapman. The possibility of an upward discharge from the top of the thunder-cloud to the ionosphere as causing a slow disturbance which may be identified with the final slow field change has also been suggested by Appleton and Chapman (1937).

It has now been established that during both the periods of slow and rapid changes of moment there are superposed minor pulsations, the most prominent pulsations occurring during the short period of rapid field change. As originally suggested by Simpson (1924), electrical oscillations are possible when the resistance of a highly ionized lightning channel becomes less than the critical value. These oscillations in the channel are actually set up and superimposed upon the main current without reversing the direction of the main current.

It is evident that the wave-form of the field change at a distance from the discharge column resulting from the destruction of a cloud moment is determined by the individual wave-forms of the *electrostatic*, *induction* and the *radiation* field changes and by their amplitudes at that distance. For extremely small distances the electrostatic field is predominant, and the wave-form of the field is *aperiodic*. At distances where the electrostatic and inductive field changes are still more prominent than the radiation field, the wave-form is still of aperiodic type. At large distances where the radiation field is predominant, the wave-form is of *quasi-periodic* type. The oscillograms of the received atmospheric pulses from different distances have shown the evolution of these wave-forms.

If we accept Schonland's idea as to the discharge mechanism of the lightning stroke, it is expected that a low frequency component will precede and follow the high-frequency component in regular time sequence, the leader streamer and the last portion of the return streamer giving rise to the initial and final slow components and the extremely quick earlier part of the return streamer producing the field change of comparatively higher frequency. The high-frequency component preceded and followed by a slow component has been observed a number of times during the present investigation. This is however expected for some moderate distance of the lightning stroke. When a slow component has been observed ahead of the high-frequency group, it is to be attributed to the leader streamer which is of comparatively long duration and precedes a return streamer. These oscillograms are evidently due to lightning discharges at moderate

distances. It is to be noted that the slow pulse appearing before the *quasi*-periodic pulse which represents the high frequency component is not due to the electrostatic effect of the slow field change associated with the leader; for it is certain that at distances where the *quasi*-periodic pulse appears, the electrostatic field change does not produce any appreciable effect. For longer distances the slow component arising out of the leader streamer would appear as a "tail" due to the slightly higher group velocity of the high frequency group in *dispersive* transmission through the ionosphere. The slow component associated with the last portion of the return streamer would, of course, appear later than the high frequency group which is associated with the extremely rapid earlier part of the return streamer.

The "precursors" observed by other workers have also been noticed in the present investigation (see figs. 6 and 7 marked "P"). The precursor was usually $1/5$ th of the main oscillatory component in amplitude and preceded the latter by an interval of 5.19 msecs. Their *quasi*-frequencies were comparatively higher than those of the oscillatory part. As suggested by Lutkin, Herd and Watson Watt (1937), the precursors are radiations during the minor discontinuities in the initial slow change of the field which have already been associated with the various branchings of the first leader streamer identified by Schonland, Malan and Collens.

There was evidence of multiple reflection between the ionosphere and the earth in a few oscillograms. The wavy border (duration of a cycle being about 0.58 msec.) on some of the slow components clearly indicated the reflection mechanism.

The slow periodic components (type VI., see fig. 7) which were observed as many as 48 times consisted of two or three half-cycles, the frequency ranging between 130 and 1265 cycles/sec. The wavy border on them is a noticeable feature. These are perhaps associated with electric impulses produced by the stoppage of meteors in the ionosphere. This suggestion is however not based on any experimental evidence.

Let us now consider the sign of the field change due to the atmospherics. According to Simpson's theory of thunderstorms, the air in a thundercloud is charged with negative electricity, thus giving a negative electric moment to the thundercloud. The destruction of the negative thundercloud moment will naturally produce a positive change of field, since on the destruction of the negative thundercloud moment, the bound positive charge on the aerial will be released and pass to earth down the earth lead. Thus the A^+ and Q^+ type of atmospherics corresponds to a positive field change which is associated with the cloud-to-earth discharge of negative charge on the thundercloud. It should be mentioned that in the proximity of the discharge channel C. T. R. Wilson (1920) found the positive discharges to be 1.56 times as frequent as negative discharges. The work of Appleton, Watson Watt and Herd (1926) on distant atmospherics revealed that the negative changes were at least 1.7 times as frequent as the positive changes. The latter results were explained by assigning a bi-polar nature

of cloud with positive charge uppermost and the negative charge underneath, the resultant moment being positive. It was shown that there was a initial distance within which the field of the bi-polar cloud would be negative and beyond which it would be a positive value. Thus both the positive and negative field changes observed could be explained. In the present investigation where a preponderance of positive sign was observed, the results indicated either the destruction of a negative thundercloud moment or that of a bi-polar cloud explained above for distances greater than the critical distance.

§6. SUMMARY.

With an equipment consisting of a damped horizontal aerial, the time-constant of which was small as compared with the duration of the atmospheric pulses, a resistance-capacity-coupled amplifier with constant voltage-gain over a range of 90 cycles/sec. to 12 kc./s. and a cathode ray oscillograph with linear time base, as many as 433 wave-forms were recorded. These represented the voltage variations across a suitable condenser placed in series with the aerial.

In addition to the usual aperiodic and *quasi*-periodic pulses, the aperiodic or *quasi*-periodic pulses followed by slow components, the precursors and the "repeater" pulses due to multiple reflection between the earth and the ionosphere, the following observations were made :

- (i) Aperiodic or *quasi*-periodic pulse preceded by a slow component.
- (ii) Aperiodic or *quasi*-periodic pulse preceded and followed by slow components.
- (iii) Slow periodic components of average duration of 3.01 msec. per cycle.

Accepting Schonland's view of the discharge mechanism of a lightning stroke, the *leader* streamer and the last portion of the *return* streamer characterized by a continuance of the channel luminosity after it has entered the cloud are considered as giving rise to the initial and final slow components respectively. Since the quick part of the return streamer is accepted as the origin of the high frequency component, it is expected that at some moderate distance, the aperiodic or *quasi*-periodic pulse which represents the high frequency component should have slow components on *either* side. The aperiodic or *quasi*-periodic pulse preceded and followed by slow components, which were at times observed can thus be explained. The slow component appearing ahead of the high frequency group in some oscillograms represents the initial slow field change attributed to the *leader* streamer. Such oscillograms are caused by lightning discharges at moderate distances. For longer distances, the slow component arising out of the *leader* streamer would appear as a "tail" due to the slightly higher group velocity of the high frequency component in *dispersive* transmission through the ionosphere. The slow component arising out of the last part of the *return* streamers would, of course, appear after the high frequency group which is associated with the extremely rapid earlier part of the return streamer.

The observed two or three cycles of slow components may be associated with the electrical impulses produced by the stoppage of meteors in the ionosphere.

Our sincere thanks are due to Mr. N. C. Dhar, M.Sc., Meteorological Assistant, India Meteorological Dept. for technical assistance during the investigation.

REFERENCES.

- APPLETON, and CHAPMAN, 1937, *Proc. Roy. Soc. A*, **158**, 1.
 APPLETON, WATSON WATT, and HERD, 1926, *Proc. Roy. Soc. A*, **111**, 615, 654.
 LUTKIN, HERD, and WATSON WATT, 1937, *Proc. Roy. Soc. A*, **162**, 267.
 MCEACHRON, and MCMORRIS, 1937, *Gen. Elec. Rev.*, **39**, 487.
 SCHONLAND, 1938, *Proc. Roy. Soc. A*, **164**, 132.
 SCHONLAND, MALAN, and COLLENS, 1935, *Proc. Roy. Soc. A*, **152**, 595.
 SIMPSON, 1926, *Proc. Roy. Soc. A*, **111**, 56 ; 1929, *J. Inst. Elec. Eng.*, **67**, 1275.
 WALTER, 1903, *Ann. der Physik*, **10**, 393 ; 1935, *Ibid.*, **52**, 421.
 WILSON, C. T. R., 1920, *Phil. Trans. Roy. Soc. A*, **221**, 75.

CIV. *Electrodynamics in a Rotating Frame of Reference.*

By M. G. TROCHERIS *.

[Received July 26, 1949.]

INTRODUCTION.

IN this paper we investigate the form of electrodynamics in a rotating frame of reference and its application to a few problems involving the electromagnetic field of a uniformly rotating body.

I. DEFINITION OF THE TRANSFORMATION.

It is impossible in general relativity to define in a unique way what is a frame of reference in uniform rotation about an axis. Various definitions have been proposed in each of which a particular property of the classical rotation is retained. A commonly used transformation (Eddington 1921, Langevin 1921) is the classical change of Cartesian coordinates between two systems xyz and $x'y'z'$ in uniform relative rotation about their common z -axis :

$$\left\{ \begin{array}{l} x=x' \cos \omega t' + y' \sin \omega t', \\ y=y' \sin \omega t' - x' \cos \omega t', \\ z=z', \\ t=t', \end{array} \right. \quad \text{OR} \quad \left\{ \begin{array}{l} r=r', \\ \theta=\theta' - \omega t', \\ z=z', \\ t=t'. \end{array} \right.$$

Let us consider this transformation in a finite region and make the axis of rotation go away to infinity while ω tends to zero in such a way

* Communicated by Professor Rosenfeld.

as to give a finite limit v for ωr . The uniform rotation tends to a uniform translation with velocity v and the transformation tends to a Galilean transformation. It would be desirable to have a transformation which in this limit tends to a Lorentz transformation.

The Transformation in First Order Approximation.

The first order approximation in ω means that we restrict ourselves to distances r of the axis of rotation small enough to neglect $(\omega r/c)^2$. We attempt to take as infinitesimal transformation in the neighbourhood of a point P the Lorentz transformation with velocity $\mathbf{v} = \boldsymbol{\omega} \wedge \mathbf{r}$. It is to a first approximation in v/c :

$$rd\theta = r'd\theta' - \omega r' dt',$$

$$dt = dt' - \frac{\omega r'^2}{c^2} d\theta',$$

$$dr = dr',$$

$$dz = dz',$$

where $r'\theta'z'$ and $r\theta z$ are the cylindrical coordinates of P around the axis of rotation in the system at rest and in the rotating system respectively. This infinitesimal transformation is not integrable, because dt is not a perfect differential of a function of $r'\theta't'$. This means that two such infinitesimal transformations taken for distances r and $r + \Delta r$ from the axis of rotation contradict each other in their common range between r and $r + \Delta r$. As the expression of dt' contains a term $-\frac{\omega r'^2}{c^2} d\theta'$ in $d\theta'$ depending on r' it must contain a term in dr' depending on θ' such that dt is the total differential of

$$t = t' - \frac{\omega r'^2}{c^2} (\theta' - \theta'_0),$$

$$dt = dt' - \frac{\omega r'^2}{c^2} d\theta' - \frac{2\omega r'}{c^2} (\theta' - \theta'_0) dr'.$$

We shall choose coordinates in both systems so that the constant θ'_0 vanishes.

The simplest transformation to a first order approximation in ω is therefore with the corresponding infinitesimal transformation

$$\left. \begin{aligned} r &= r', & dr &= dr' & dz &= dz', \\ \theta &= \theta' - \omega t', & d\theta &= d\theta' - \omega dt', \\ t &= t' - \frac{\omega r'^2}{c^2} \theta', & dt &= dt' - \frac{\omega r'^2}{c^2} d\theta' - \frac{2\omega r'}{c^2} \theta' dr'. \end{aligned} \right\} \dots \dots (1)$$

The converse formulæ are obtained by changing ω into $-\omega$.

Transformation for all Values of ω .

To determine the transformation for all values of ω we must impose a new condition on it. We choose to retain the group property of the classical uniform rotations about an axis: if two frames of reference are both in uniform rotation about the same axis then they are in uniform rotation with respect to each other. There is only one transformation satisfying this requirement and reducing to our transformation in the first approximation in ω . It is determined as the one parameter group of transformations in $r\theta zt$ corresponding to the infinitesimal transformation

$$\begin{aligned}\delta r &= \delta z = 0, \\ \delta \theta &= t \delta \omega, \\ \delta t &= \frac{r^2 \theta}{c^2} \delta \omega\end{aligned}$$

and it is found to be

$$\left. \begin{aligned}\theta' &= \theta \operatorname{ch} \frac{\omega r}{c} + \frac{ct}{r} \operatorname{sh} \frac{\omega r}{c}, \\ t' &= t \operatorname{ch} \frac{\omega r}{c} + \frac{r\theta}{c} \operatorname{sh} \frac{\omega r}{c}, \\ r' &= r, \\ z' &= z,\end{aligned} \right\} \dots \dots \dots (2)$$

$$\left. \begin{aligned}\theta &= \theta' \operatorname{ch} \frac{\omega r'}{c} - \frac{ct'}{r'} \operatorname{sh} \frac{\omega r'}{c}, \\ t &= t' \operatorname{ch} \frac{\omega r'}{c} - \frac{r'\theta'}{c} \operatorname{sh} \frac{\omega r'}{c}, \\ r &= r', \\ z &= z',\end{aligned} \right\}, \dots \dots \dots (3)$$

where $r'\theta'z't'$ and $r\theta zt$ are respectively the coordinates in the rest system and in the rotating system. One can easily verify that, if ω and ω' are the angular velocities of S' with respect to S and of S'' with respect to S' , $\omega'' = \omega + \omega'$ is the angular velocity of S'' with respect to S . Let us show that a fixed point of the rotating system has never a velocity larger than c for the fixed observer. The coordinates $r'\theta'z't'$ of such a point in the rest system are given by the formulæ (2), where $r\theta z$ are constant and t varies. The point therefore moves round a circle $r' = r$ with an angle θ' depending on the time t' according to

$$\theta' = \theta'_0 + t' \frac{c}{r} \operatorname{th} \frac{\omega r}{c}.$$

Its velocity is

$$v' = r' \frac{d\theta'}{dt'} = c \operatorname{th} \frac{\omega r}{c},$$

which is always smaller than c and tends to c when r tends to infinity.

If we put $v/c = \text{th } \omega r/c$ the transformation can be written

$$r'\theta' = \frac{1}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}} (r\theta + vt) \quad r' = r,$$

$$t' = \frac{1}{\sqrt{\left(1 - \frac{v^2}{c^2}\right)}} \left(t + \frac{vr\theta}{c^2}\right) \quad z' = z.$$

This reduces to a Lorentz transformation far from the axis of rotation insofar as $\omega r/c$ can be regarded as a constant in the region considered. It tends rigorously to a Lorentz transformation in a finite region when the axis of rotation goes away to infinity and $\omega r/c$ tends to a finite constant. The velocity of the Lorentz transformation is not the limit of ωr but of $c \text{ th } \omega r/c$.

II. KINEMATICS IN THE ROTATING SYSTEM.

We shall have to consider in the following the time taken by the light to go along various paths. We suppose that it is possible to realize such paths by suitable devices of mirrors.

Let us first consider the light going round a "parallel" $r = \text{const.}$, $z = \text{const.}$ On such a circle, where r is constant, the transformation, taken in the form (4) is mathematically identical with a Lorentz transformation on $r\theta$ and t . Therefore the composition of the velocities round such a circle is the same as that of linear velocities, and in the rotating system the light goes round such a circle with velocity c in both directions.

A consequence is that two clocks A, B on the same parallel can be adjusted in the following way: a light signal is transmitted for A at t_1 , goes round the circle, is reflected by B and received back in A at t_2 . The two clocks are adjusted if the reflection occurs on the clock B when it shows $(t_1 + t_2)/2$.

The point events time t at A and B in the rotating system have, for the fixed observer, time coordinates which differ by

$$(\theta_B - \theta_A) \frac{r}{c} \text{sh } \frac{\omega r}{c}.$$

This difference depends on the exact value chosen for θ_A and θ_B . For example, consider two clocks A and B at the same point in the rotating system showing the same time for both rotating and fixed observer. Let us rotate B clockwise by 2π round the axis of rotation in the rotating system. When B coincides again with A the rotating observer will still read the same time by both clocks, but, for the fixed observer, B will

be slow with respect to A by $\Delta t' = 2\pi \frac{r}{c} \text{sh } \frac{\omega r}{c}$.

The two main points about the exact determination of θ in the rotating system are :

(1) To define the exact value of θ for a clock in the rotating system is equivalent to defining the time shown by this clock for the fixed observer.

(2) A definite value of θ cannot be ascribed to an abstract point in the rotating system but only to a material support whose coordinate θ is bound to vary continuously in its displacements.

The same can be said about the value of θ' for a clock in the rest system and the time it shows for the rotating observer.

In the following applications we shall use the transformation to a first approximation in the angular velocity. To this approximation the interval between two neighbouring point events in the rotating system is given by *

$$ds^2 = dr^2 + r^2 d\theta^2 + dz^2 - c^2 dt^2 - 4\omega r \theta dt dr = dl^2 - c^2 dt^2 - 4\omega r \theta dt dr. \quad (5)$$

The time taken by light to travel between the two points r, θ, z and $r+dr, \theta+d\theta, z+dz$ is given by the condition $ds=0$ and found to be

$$dt = \frac{dl}{c} - \frac{2\omega r \theta}{c^2} dr. \quad (6)$$

Theory of Sagnac's Experiment.

In Sagnac's (1913) experiment a beam of light is divided at a point A into two beams going in opposite directions around a closed circuit which is realized by a suitable device of mirrors. The two beams are caused to interfere when they meet again at A. The mirrors are supported by a platform which can be rotated. When the platform is set in rotation with angular velocity ω a phase shift is observed showing that the two beams return to A after two times of travel $t'_1 - t'_0$ and $t'_2 - t'_0$, which differ by

$$t'_1 - t'_2 = \frac{4\omega S}{c^2},$$

where S is the area enclosed by the circuit followed by the light or rather its projection on a plane perpendicular to the axis of rotation.

Let us suppose that this circuit is a parallel. In the rotating system the light takes the same time to go round the circle in both directions and the two rays which left A at the same time t_0 return to A at the same time $t_1 = t_2$. However, their arrivals back at A are two distinct point events with different coordinates θ :

Ray 1	$r_1 = R,$	$\theta_1 = \theta_0 + 2\pi,$	$t_1 = t_0 + 2\pi R/c.$
Ray 2	$r_2 = R,$	$\theta_2 = \theta_0 - 2\pi,$	$t_2 = t_1.$

* The corresponding Riemannian space has a vanishing curvature and there the g^{ik} satisfy the equations of gravitation for empty space.

They will therefore have for the fixed observer two different time-coordinates t'_1, t'_2 , such that

$$t'_1 - t'_2 = \frac{\omega R^2}{c^2} (\theta_1 - \theta_2) = 4\pi \frac{\omega R^2}{c^2} = \frac{4\omega S}{c^2},$$

which is Sagnac's result. With the help of (6) the result is easily found to hold for any shape of the circuit followed by the light.

Application to the Earth.

In a Gallilean frame of reference (in uniform translation with respect to the stars) simultaneity can be defined between two points by assuming that the time taken by the light to go from one to the other is the same in both directions along the same path. But this is not possible for any two points in a rotating system and any path between them. It would lead in the case of the earth to discrepancies of the order 10^{-8} to 10^{-7} seconds. Beyond this accuracy a precise definition of the time on the earth would be needed. A simple one is provided by our transformation. With it two clocks which are adjusted for an observer rotating with the earth would no longer be adjusted for an observer who does not rotate with the earth. For two clocks whose longitudes differ by 180° the difference is $\pi\omega R^2/c^2 \sim 10^{-7}$ sec. The clock which is to the west will be slow with respect to that which is to the east. It would be the opposite for two clocks adjusted for the fixed observer and read by the rotating observer.

III. ELECTRODYNAMICS IN THE ROTATING SYSTEM.

We now give the laws of electrodynamics as they are experienced by the rotating observer to a first order approximation in ω . They are obtained from their general covariant form as given by general relativity by taking the fundamental form (5).

Transformation of the Field Equations.

The field in the rotating system is defined relative to the rotating system by two vectors \mathbf{E}, \mathbf{H} the electric and magnetic field strengths. The corresponding vectors in the rest system are

$$\left. \begin{aligned} \mathbf{E}' &= \mathbf{E} - \frac{1}{c} (\mathbf{v} \wedge \mathbf{H}), \\ \mathbf{H}' &= \mathbf{H} + \frac{1}{c} (\mathbf{v} \wedge \mathbf{E}) + \frac{2\omega\theta}{c} (\mathbf{r} \wedge \mathbf{E}), \end{aligned} \right\} \mathbf{v} = \omega \wedge \mathbf{r}. \quad \dots (7)$$

In material bodies the field is defined by the four usual vectors. $\mathbf{E}, \mathbf{D}, \mathbf{B}, \mathbf{H}$. The above formulæ hold separately for \mathbf{E}, \mathbf{B} and \mathbf{D}, \mathbf{H} .

The charge and current densities transform according to

$$\mathbf{j}' = \mathbf{j} + (\omega \wedge \mathbf{r})\rho, \quad \rho' = \rho + \frac{1}{c^2} (\omega \wedge \mathbf{r}) \cdot \mathbf{j} + \frac{2\omega\theta}{c} \mathbf{j} \cdot \mathbf{r}. \quad \dots (8)$$

The converse formulæ are obtained by changing ω into $-\omega$.

Laws of Electrodynamics in the Rotating System.

Maxwell's equations in material bodies are replaced by the following equations:

$$\left. \begin{aligned} \operatorname{curl} \left[\mathbf{H} + \frac{2\omega\theta}{c} \mathbf{r} \wedge \mathbf{D} \right] - \frac{1}{c} \frac{\partial}{\partial t} \left[\mathbf{D} + \frac{2\omega\theta}{c} \mathbf{r} \wedge \mathbf{H} \right] &= \frac{4\pi}{c} \mathbf{j}, \\ \operatorname{div} \left[\mathbf{D} + \frac{2\omega\theta}{c} \mathbf{r} \wedge \mathbf{H} \right] &= 4\pi\rho, \\ \operatorname{curl} \mathbf{E} + \frac{1}{c} \frac{\partial \mathbf{B}}{\partial t} &= 0, \\ \operatorname{div} \mathbf{B} &= 0. \end{aligned} \right\} \quad \dots \quad (9)$$

The second set expresses the fact that \mathbf{E} , \mathbf{B} are derived from potentials and it has the usual form. The relations $\mathbf{B}=\mu\mathbf{H}$, $\mathbf{D}=\epsilon\mathbf{E}$ inside a body are no longer valid in the rotating system. At a point where the body moves with velocity \mathbf{u} in the rotating system they have to be replaced by

$$\begin{aligned} \mathbf{D} + \frac{1}{c} \mathbf{u} \wedge \mathbf{H} &= \epsilon (\mathbf{E} + \frac{1}{c} \mathbf{u} \wedge \mathbf{B}) \\ \mathbf{B} - \frac{1}{c} \mathbf{u} \wedge \mathbf{E} + \frac{2\omega\theta}{c^2} (\mathbf{u} \cdot \mathbf{r}) \mathbf{B} + \frac{2\omega\theta}{c} \mathbf{r} \wedge \mathbf{E} &= \mu \left[\mathbf{H} - \frac{1}{c} \mathbf{u} \wedge \mathbf{D} + \frac{2\omega\theta}{c^2} (\mathbf{u} \cdot \mathbf{r}) \mathbf{H} + \frac{2\omega\theta}{c} \mathbf{r} \wedge \mathbf{D} \right]. \end{aligned}$$

The equations of the field in vacuum are obtained by putting $\mathbf{D}=\mathbf{E}$, $\mathbf{B}=\mathbf{H}$, $\rho=\mathbf{j}=0$. They can be derived from the lagrangian

$$L = E^2 - H^2 - \frac{4\omega\theta}{c} \mathbf{H} \cdot (\mathbf{r} \wedge \mathbf{E}).$$

The definition of charge and current densities in terms of moving particles is unchanged to a first approximation in the velocity of the particles.

The continuity equation has the usual form

$$\operatorname{div} \mathbf{j} + \frac{\partial \rho}{\partial t} = 0.$$

Rather complicated expressions give the retarded potentials due to a given distribution of charge and current

$$\left. \begin{aligned} \mathbf{A}(x_0, y_0, z_0, t_0) &= \frac{1}{c} \int \mathbf{J}(x, y, z, t_0 + \Delta t) \frac{dx dy dz}{R}, \\ \phi(x_0, y_0, z_0, t_0) &= \int \kappa(x, y, z, t_0 + \Delta t) \frac{dx dy dz}{R}, \end{aligned} \right\} \quad \dots \quad (10)$$

$$\begin{aligned} R &= \sqrt{\{(x-x_0)^2 + (y-y_0)^2 + (z-z_0)^2\}} \\ \Delta t &= -\frac{R}{c} - \frac{\omega}{c^2} [xy_0 - yx_0 + r^2\theta - r_0^2\theta_0], \end{aligned}$$

\mathbf{J} and κ are linear combinations of \mathbf{j} and ρ , depending on the point x, y, z , where \mathbf{j}, ρ are considered and also on the point $x_0 y_0 z_0$, where the potentials are sought. They might be called "apparent" charge and current densities at the point xyz for the observer at the point $x_0 y_0 z_0$. Their expressions are

$$\left. \begin{aligned} \mathbf{J} &= \mathbf{j} - \frac{R}{c} \boldsymbol{\omega} \wedge \mathbf{j} + \boldsymbol{\omega} \wedge (\mathbf{r} - \mathbf{r}_0) \rho - 2\omega \theta_0 r_0 \rho, \\ \kappa &= \rho + \frac{2\omega \theta}{c^2} \mathbf{r} \cdot \mathbf{j} + \frac{1}{c^2} \boldsymbol{\omega} \cdot [(\mathbf{r} + \mathbf{r}_0) \wedge \mathbf{j}]. \end{aligned} \right\} \dots \dots (11)$$

A consequence is that a distribution of charges without any current can give rise to a magnetic field.

IV. APPLICATIONS.

Before treating a few applications it is interesting to notice that our transformation cannot yield straight away the electromagnetic field of rotating bodies as a Lorentz transformation does for translating bodies. If one knows the field of a certain body, for instance a magnet, when it is at rest, one can derive the field it produces, when in uniform translation. It is obtained by a Lorentz transformation of the field of the translating magnet in a frame of reference moving with it. But a Lorentz transformation does not by itself say what this latter field is for it cannot say what may happen to the atomic structure of a magnet when it is set in translation. However, the principle of relativity states that no experiment on the field of a magnet can detect an absolute uniform translation and therefore that the field of the moving magnet for the moving observer is the same as if both magnet and observer are at rest. In the case of rotation there is no such principle. An absolute rotation with respect to the stars can be detected by suitable experiments. Anyhow, it can easily be realized that a mere transformation cannot tell what a body becomes when it is set in rotation; for instance, the centrifugal forces will distort it according to its particular elastic properties.

Field of a Rotating Spherical Condenser.

The following paradox was mentioned and solved by Schiff (1939). No field is observed outside a charged spherical condenser if it is at rest. If it is set in rotation about a diameter the field of a magnetic dipole is seen outside. But an observer rotating about the condenser at rest cannot observe any field because if there is no field for one observer there is no field for any observer. Why does he not observe the same magnetic field as the observer at rest when he sees the condenser rotating? There is no field for the observer rotating about the condenser because the laws of electrodynamics in the rotating system are different from those in the rest system and such as to give no field. The calculation can be made with our equations. The electric field can be shown to vanish outside the condenser just as in the rest system. The charges

are seen to move with velocity $-\boldsymbol{\omega} \wedge \mathbf{r}$ and give rise to a current density $\mathbf{j} = -\boldsymbol{\omega} \wedge \mathbf{r} \rho$. But the magnetic field has to be derived from the apparent current density given by (11). One can show that it reduces in this case to

$$\mathbf{J} = \mathbf{j} + \boldsymbol{\omega} \wedge \mathbf{r} \rho,$$

and therefore vanishes and yields no magnetic field.

Field of a Rotating Current.

Let us first consider the simpler case of a translating current. Suppose that we observe a uniformly translating conductor through which a current flows. This conductor may be a complete circuit or part of a circuit. According to relativity we see in the conductor a certain charge density as well as a current density, whereas an observer moving with the conductor would see only a current density. This is a purely relativistic effect and is merely due to the different ways in which the motion of charged particles is observed. The effect can be seen from the relativistic conduction equations which give the current and charge densities raised by an electromagnetic field in a uniformly translating conductor. These equations are in a first approximation in the velocity \mathbf{u}' of the conductor

$$\left. \begin{aligned} \mathbf{j}'_c &= \sigma \left(\mathbf{E}' + \frac{1}{c} \mathbf{u}' \wedge \mathbf{B}' \right), \\ \rho'_c &= \frac{\sigma}{c^2} (\mathbf{u}' \cdot \mathbf{E}'), \quad \rho'_c = \frac{1}{c^2} (\mathbf{u}' \cdot \mathbf{j}'_c). \end{aligned} \right\} \dots \dots \dots (12)$$

It cannot be proved by means of our transformation that these expressions for ρ'_c and \mathbf{j}'_c hold, in the same approximation, for each element of a rotating current considered separately with its own velocity $\mathbf{u}' = \boldsymbol{\omega} \wedge \mathbf{r}$. But it seems a natural assumption. With this assumption the expression for the conduction current and charge densities in the rotating system are

$$\begin{aligned} \mathbf{j}_c &= \sigma \left[\mathbf{E} + \frac{1}{c} \mathbf{u} \wedge \mathbf{B} + \frac{2\omega\theta}{c^2} (\mathbf{u} \cdot \mathbf{E}) \mathbf{r} \right], \\ \rho_c &= \sigma \left[\frac{1}{c^2} (\mathbf{u} \cdot \mathbf{E}) - \frac{2\omega\theta}{c^2} \mathbf{r} \cdot \left(\mathbf{E} + \frac{1}{c} \mathbf{u} \wedge \mathbf{B} \right) \right], \end{aligned}$$

where u is the velocity of the conductor at the point considered. If \mathbf{u} is the same order of magnitude as ωr they satisfy the relation

$$\rho_c = \frac{1}{c^2} (\mathbf{u} \cdot \mathbf{j}_c).$$

A solenoid rotating about its own axis, is seen by an observer at rest to be uniformly charged with a density

$$\rho'_c = \frac{\omega r}{c^2} j'_c,$$

which is positive if the rotation is in the direction of the current.

j'_c and ρ'_c are not necessarily the total current and charge densities which would be observed. For there can be surface charges on the rotating conductor. These surface charges could screen the effect of the charge density ρ'_c and it is interesting to know how they vary for a given current. In the first order approximation in ω they produce only an electric field. For a given j'_c the electric field inside the conductor is determined by (12) and cannot vary. Therefore the surface charges can only vary by an equilibrium distribution which gives no field inside and are completely determined if the electric potential is fixed at one point of the surface.

Electric Field of a Rotating Cylindrical Magnet.

Let us consider a magnet which has cylindrical symmetry, for instance a cylindrical magnet. If it is rotating about its axis it produces an electric field outside and a "unipolar induction" e.m.f. can be observed round a circuit including the rotating magnet itself, and a fixed wire between two brushes touching the magnet one near the pole and one near the equator (Swann 1920, Tate 1922). The e.m.f. observed is the same as when the magnet is at rest and the wire rotating about it. It seems to depend only on the relative rotation of the magnet and the wire. It is interesting to investigate how far the problem of unipolar induction for an observer rotating with the magnet is the same as the simple problem of induction in a wire rotating about a fixed magnet. We shall first consider the electric field and the e.m.f. for an observer at rest and then transform them to the rotating system.

The electric field of a rotating conducting magnet can be calculated in the following way (Swann 1920). The free electrons inside the magnet are acted on by the Lorentz force $\mathbf{E}' + \frac{1}{c} \mathbf{v} \wedge \mathbf{B}'$. One looks for a state of equilibrium in which they are stationary relative to the magnet. In a first approximation in ω the condition for this is that the Lorentz force vanishes:

$$\mathbf{E}' = -\frac{1}{c} (\mathbf{v} \wedge \mathbf{B}'). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

This electric field inside the magnet is created by charges which can be calculated and are determined only up to an equilibrium surface distribution which does not change the field inside. As in the case of the rotating solenoid the surface charges are completely determined if the electric potential is fixed at one point of the surface. The external electric field is due to both the internal and surface charges; it is not equal to $-\frac{1}{c} \mathbf{v} \wedge \mathbf{H}'$, and one has only

$$\int_A^B \mathbf{E}' \cdot d\mathbf{s}' = \int_A^B -\frac{1}{c} (\mathbf{v} \wedge \mathbf{H}') \cdot d\mathbf{s}',$$

where $d\mathbf{s}'$ is the line element and the integrals are taken along any path outside the magnet between two points A, B of its surface.

It is not clear at first sight how, with such an electric field, there can be a unipolar induction e.m.f. Let us first see why there is an induction e.m.f. when the wire is rotating about the fixed magnet, even if it is in a meridian plane in which case there is no variation of flux (fig. 1).

There is an e.m.f. in a circuit if work is done by the electromagnetic force, or more precisely by its component along the circuit, when an electron goes once round the complete circuit. The current is supposed to be small enough for its field to be negligible compared with that of the magnet. The work of the electromagnetic force vanishes when the

Fig. 1.

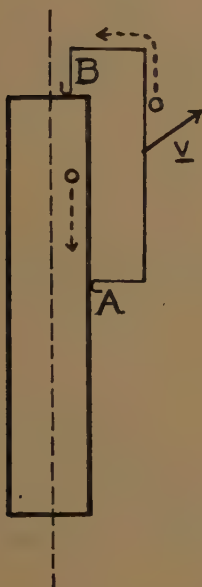
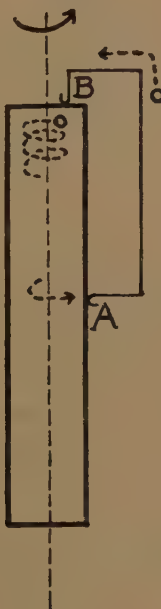


Fig. 2.



electron goes from B to A inside the magnet. When it goes from A to B through the wire the component of the magnetic force along the wire is equal to that of $\frac{1}{c} \mathbf{v} \wedge \mathbf{H}'$, where \mathbf{v} is the velocity of the wire due to its rotation and the work done is

$$\frac{e}{c} \int_A^B (\mathbf{v} \wedge \mathbf{H}') \cdot d\mathbf{s}' = e\mathcal{E}',$$

the integral being taken along the wire. This determines the expression for the e.m.f. \mathcal{E}' . The unipolar induction e.m.f. can be found in the same way (fig. 2). The work done when the electron moves slowly through the wire from A to B reduces to that of the electric field

$$e(V'_A - V'_B) = e \int_A^B \mathbf{E}' \cdot d\mathbf{s}' = -\frac{e}{c} \int_A^B (\mathbf{v} \wedge \mathbf{H}') \cdot d\mathbf{s}' = e\mathcal{E}'. \quad \dots (14)$$

When the electron arrives in the magnet at B it is taken into the rotation. Small extra charges, which were accumulated when the current was switched on, drive it slowly inside the magnet and its spirals from B to A.

The total force $e\left(\mathbf{E}' + \frac{1}{c} \mathbf{v} \wedge \mathbf{B}'\right)$ acting on it vanishes according to (13). Thus the rotation of the magnet enables the electron to go back from B to A inside the magnet without any work, and so the e.m.f. along the complete circuit is given by (14). This result is often expressed by saying that the lines of force rotate with the magnet and cut the wire with velocity \mathbf{v} so that the e.m.f. is the same as if the wire was cutting the lines of force with velocity $-\mathbf{v}$.

It follows from the condition (13) that there is no electric field inside the magnet for an observer rotating with it, just as if both magnet and observer were at rest. But outside the magnet the rotating observer sees an electric field \mathbf{E} which satisfies

$$\text{curl } \mathbf{E} = 0, \quad \text{div} \left(\mathbf{E} + \frac{2\omega\theta}{c} \mathbf{r} \wedge \mathbf{H} \right) = 0.$$

By taking into account that \mathbf{E}' , \mathbf{B}' are constant in time and have a symmetry of revolution in the rest system, one can show that \mathbf{E} , \mathbf{B} are constant in time in the rotating system. Further, \mathbf{E} , \mathbf{B} can be derived from potentials ϕ , \mathbf{A} according to

$$\mathbf{E} = -\text{grad } \phi, \quad \mathbf{B} = \text{curl } \mathbf{A}.$$

ϕ is constant inside the magnet and continuous on the surface. It is therefore constant on the surface. The quantities involved in the calculation of the e.m.f. are of the first order in ω , and the relations between them are, in our approximation, the same as in the rest system. Therefore the electric field outside which has a potential constant on the surface does not give any contribution to the e.m.f. which turns out to be the same as for a stationary magnet and a rotating wire, as measured by an observer at rest. We can thus conclude that the e.m.f. only depends on the relative rotation of the magnet and the wire. The electric field inside the magnet depends only on the relative rotation of the magnet and the observer, but the electric field outside depends on their absolute rotation.

I wish to express my thanks to Professor Blackett and Professor Rosenfeld for their interest in this work and many stimulating discussions.

REFERENCES.

- EDDINGTON, 1921, *Mathematical Theory of Relativity* (Cambridge).
 LANGEVIN, 1921, "Theory of Sagnac's Experiment", *C.R.*, **173**, 831.
 SAGNAC, 1913, *C.R.*, **157**, 708.
 SCHIFF, 1939, *Nat. Acad. Sci. Proc.*, **25**, 391.
 SWANN, 1920, *Phys. Rev.*, **15**, 365.
 TATE, 1922, *Bull. Nat. Research Council.*, Vol. 4, Part 6, p. 75.

CV. *Internal Pair Creation in Na²⁴.*

By E. R. RAE, M.A.,

Department of Natural Philosophy, University of Glasgow *.

[Received August 18, 1949.]

ABSTRACT.

A measurement has been made of the internal pair-creation coefficient of the 2.76 MeV. gamma-ray line from Na²⁴. This is found to lie very close to the curve of electric dipole transitions predicted by Jaeger and Hulme for $Z=0$ and suggests that measurement of this coefficient for high energy transitions in light nuclei may provide a useful method of determining multipole order. On the assumption that the 2.76 MeV. line is of dipole origin, an energy level diagram has been constructed for the Mg²⁴ nucleus.

INTRODUCTION.

THE internal pair-creation coefficient (I) of nuclei of charge $Z=0$ and $Z=84$ for electric dipole and quadrupole transitions was calculated by Jaeger and Hulme in 1935. Recently in a letter to *Nature*, Wang (1948) presented a curve showing the variation with energy of the coefficient for $Z=84$ and for magnetic dipole transitions. In each case the authors gave the energy spectrum of the positrons which should be produced. The energy spectra of the positrons from ThB and Radon in equilibrium with their decay products have been investigated by a number of authors (Alichanov *et al.* 1936, and later work reviewed by Latyshev 1947; Bradt *et al.* 1946) and have shown good general agreement with the theory. In particular the variation with energy of the internal pair-creation coefficient for both electric dipole and quadrupole transitions has been verified accurately by Alichanov for the above two sources, ($Z \sim 84$), over the energy range 1 to 2.62 MeV. The only absolute determination of I published however, appears to be that of Alichanov for the 2.62 MeV. line of ThC''. This agrees with the curve of electric quadrupole transitions for $Z=84$ given by Jaeger and Hulme.

No precise measurement of I seems to have been made for a light element, and so the variation of I with Z predicted by the theory has not been verified. Consequently it has been thought worth while to examine the positron spectrum of Na²⁴ which has a simple beta-decay followed by two gamma-rays in cascade of energies 1.38 and 2.76 MeV., (Wiedenbeck 1947, and see also references quoted) and to obtain an absolute measurement of the internal pair-creation coefficient for the 2.76 MeV. line. As a further

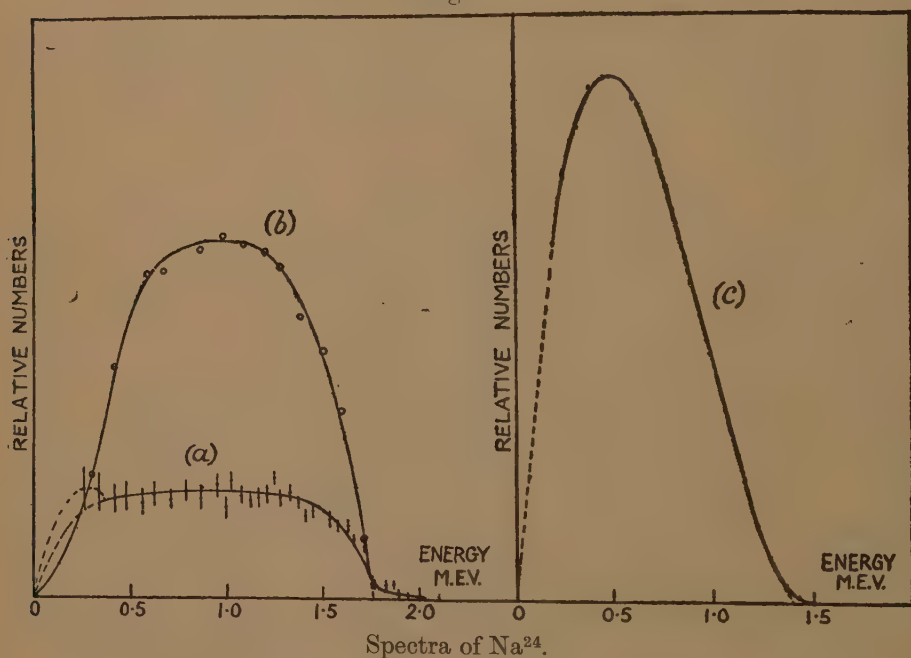
* Communicated by Professor P. I. Dee, F.R.S.

check of the theory, a direct comparison has been made between this coefficient and that of the 2.62 MeV. line of ThC'' which is known to be of electric quadrupole origin (Latyshev 1947).

EXPERIMENT.

The experiment was carried out with a semicircular focusing magnetic spectrometer of 12 cm. radius and 2 per cent. geometrical resolving power. The detector was a small double G.M. counter with a window of 0.0005" copper foil and had a background of 1.5 coincidences per minute. The Na^{24} source consisted of 100 mg. of active sodium carbonate deposited on a piece of 0.002" aluminium foil measuring 3.5 cm. by 8 mm. and had an initial strength of 10 mC. This source was mounted on a light framework

Fig. 1.

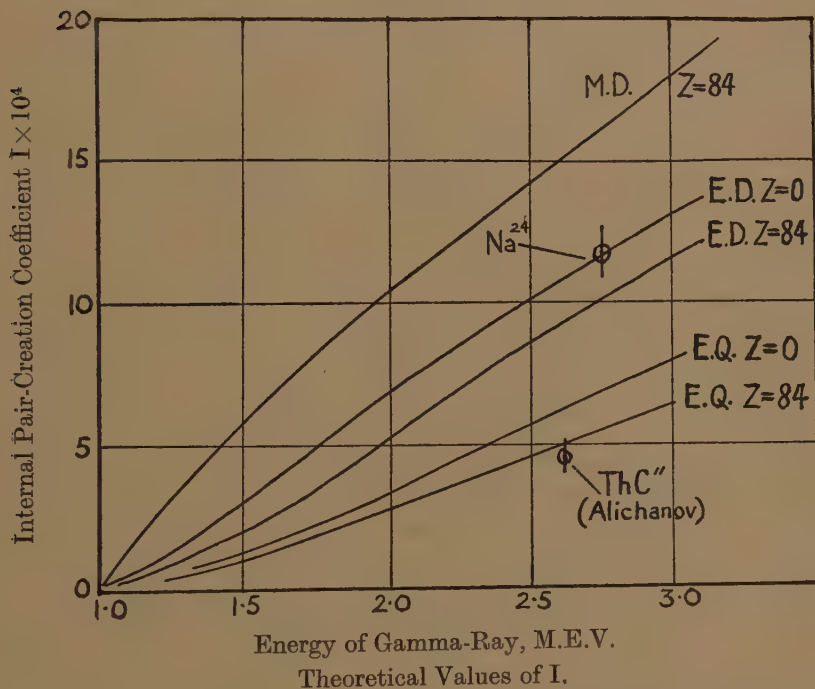


(a) Positrons from bare source. (b) Positrons from thick lead foil.
(c) Electrons.

made of the same gauge of aluminium foil, and all the walls of the vacuum-chamber were lined with aluminium to reduce to a minimum the number of stray positrons coming from these parts. The technique consisted in recording the positron spectrum while the source was as active as possible, and comparing the area under this curve with the area under the curve obtained by recording the beta-ray spectrum, the latter measurement being made after the source had aged for a few days. The curves obtained are shown in fig. 1., and the value of $I(\text{Na})$ determined from this data is $(1.16 \pm 0.10) \times 10^{-3}$. The predicted values of the coefficient are plotted in fig. 2, and it will be noticed that the above value very nearly coincides with the $Z=0$ curve for electric dipole transitions.

A further check on this result is available if the number of positrons from the sodium source is compared with the number from a ThC'' source provided the ratio of the gamma-ray intensities is known. In order to ascertain this ratio a shutter was provided on the apparatus which enabled a thick lead foil to be placed immediately in front of the source. The positron spectra of Na^{24} with and without the lead foil in place are shown in fig. 1 and a similar pair of curves was obtained with a ThC'' source. By comparing the areas under the four curves and making a

Fig. 2



small correction for the difference in energy between the two lines, a value of 2.38 ± 0.30 was obtained for the ratio $I(\text{Na})/I(\text{Th})$. Assuming that the ThC'' line is of electric quadrupole origin (Latyshev 1947) we can obtain from fig. 2 theoretical values for $I(\text{Na})/I(\text{Th})$ for the cases of the Na line being either of electric dipole or quadrupole origin. These values are 2.32 and 1.35 respectively. Hence this extrapolation from the ThC'' line also leads to the conclusion that the Na line corresponds to an electric dipole transition.

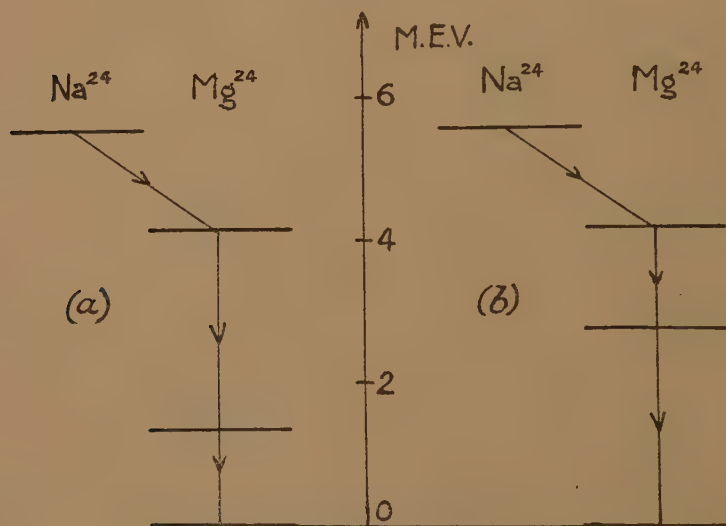
DISCUSSION.

The experiment certainly shows that the theory predicts the correct order of magnitude for the internal pair-creation coefficient of the gamma-ray investigated and that this quantity, unlike the internal conversion coefficient, does not depend markedly on Z . If we now go further and assume that the theory is substantially correct, then a

measurement of I provides an indication of the multipole order of a gamma-ray line. Furthermore as I increases with increasing energy and is almost independent of Z , this gives a method of determining multipole order which is complementary to that of measuring internal conversion coefficients, since the latter method breaks down in the case of high energy transitions in light nuclei. For this reason it is proposed to investigate the internal pair-creation in other light elements.

If we consider again the case of Na^{24} , the interpretation that the 2.76 MeV. line is of electric dipole origin enables us to establish a simple level scheme for the Mg^{24} nucleus. It has been shown (Wiedenbeck 1947) that the level scheme is given by either (a) or (b) in fig. 3. Now Pollard

Fig. 3.

Energy Level Diagram of Mg^{24} Nucleus.

and Alburger (1948) from experiments on large angle scattering of the gamma-rays of Na^{24} by Mg and Al targets, state that the transitions to the ground state of the Mg^{24} nucleus from the levels at 1.38 and 2.76 MeV., if they exist, are certainly of quadrupole or higher order. If then we accept that the 2.76 MeV. line corresponds to a dipole transition, we must reject scheme (b) and assume that scheme (a) is correct.

REFERENCES.

- ALICHANOV, A. I., ALICHANIAN, A. I., and KOSODAIEV, M. S., 1936, *J. Phys. Radium*, **7**, 163.
 BRADT, H., HALTER, J., HEINE, H. G., and SCHERRER, P., 1946, *Helv. Phys. Acta.*, **19**, 431.
 JAEGER, J. C., and HULME, H. R., 1935, *Proc. Roy. Soc.*, **148**, 708.
 LATYSHEV, G. D., 1947, *Rev. Mod. Phys.*, **19**, 132.
 POLLARD, E., and ALBURGER, D. E., 1948, *Phys. Rev.*, **74**, 926.
 WANG, M. H., 1948, *Nature Lond.*, **162**, 264.
 WIEDENBECK, M. L., 1947, *Phys. Rev.*, **72**, 429.

CVI. *Scattering of Pseudoscalar Charged Mesons by Nucleons.*—I.

By E. CORINALDESI† and G. FIELD‡,
 Dublin Institute for Advanced Studies §.

[Received August 8, 1949.]

SUMMARY.

The cross-section for the scattering of a pseudoscalar charged meson by a nucleon is evaluated by using the ordinary perturbation theory, and is expressed as a series of powers of the coupling constant. The first term of this expansion, proportional to g^4 , is finite.

The second term, proportional to g^6 , contains sums over an infinite number of intermediate states and is divergent.

In §§ 1–4 we analyse these divergences and show that they can all be interpreted as meson and nucleon mass-corrections and mesonic charge-corrections. The subtraction of these terms, by introducing appropriate counter-terms into the Hamiltonian, follows methods already developed in electrodynamics and applied to the similar problem of Compton scattering.

Our present analysis is carried out for the case of pseudoscalar coupling (g -coupling) only, with the purpose of computing also the finite terms in the sixth-order correction for the cross-section.

The finite corrections $\sim g^6$ are evaluated for the case of mesons with momenta $p \ll \mu$. The fourth-order and sixth-order contributions to the cross-section, $d\sigma_4$ and $d\sigma_6$, are evaluated and their ratio is given in terms of g . This is done in § 5.

§ 1. INTRODUCTION.

THE total Hamiltonian *density* of the pseudoscalar charged meson field is

$$\mathbf{H}_\mu = \pi^* \pi + (\text{grad } \psi^* \cdot \text{grad } \psi) + \mu^2 \psi^* \psi, \quad (1)$$

with the well-known commutation relations for the operators ψ and π (Wentzel 1949). For the proton-neutron field we have

$$\mathbf{H}_M = \Phi^* \left[\frac{1}{i} \boldsymbol{\alpha} \cdot \text{grad} + \beta M \right] \Phi, \quad (2)$$

with the usual definition of the Dirac matrices, considering the mass of

† Now at the Physics Department, Manchester University.

‡ Now at the Department of Mathematical Physics, Birmingham University.

§ Communicated by Professor W. Heitler, F.R.S.

the neutron and of the proton to be equal. Φ is an eight-component column matrix $\Phi = \begin{pmatrix} \Phi_P \\ \Phi_N \end{pmatrix}$. Putting

$$\Phi_P \equiv \begin{pmatrix} \Phi_1 \\ \vdots \\ \Phi_4 \end{pmatrix} \quad \text{and} \quad \Phi_N \equiv \begin{pmatrix} \Phi_5 \\ \vdots \\ \Phi_8 \end{pmatrix}$$

we can write for the commutation relations

$$[\Phi_e(\mathbf{x}), \Phi_\sigma^*(\mathbf{x}')]_+ = \delta_{e\sigma} \delta(\mathbf{x} - \mathbf{x}'). \quad (3)$$

In this paper we shall consider only the so-called pseudoscalar coupling for the interaction. This gives only a first-order term in the coupling constant g in the interaction Hamiltonian *density*

$$\mathbf{H}_{M\mu} = -\sqrt{(4\pi)g}(\mathbf{B}\psi + \mathbf{B}^*\psi^*), \quad (4)$$

where

$$\left. \begin{aligned} \mathbf{B} &= (\Phi^\dagger \gamma_5 \tau \Phi), & \mathbf{B}^* &= (\Phi^\dagger \gamma_5 \tau^* \Phi), \\ \Phi^\dagger &= i\Phi^* \beta, & \gamma_5 &= \gamma_1 \gamma_2 \gamma_3 \gamma_4, \\ \tau &= \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, & \tau^* &= \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}. \end{aligned} \right\} \quad (5)$$

We can also write

$$\mathbf{B} = i\Phi_P^* \beta \gamma_5 \Phi_N, \quad \mathbf{B}^* = i\Phi_N^* \beta \gamma_5 \Phi_P. \quad (6)$$

So our *Hamiltonian* is

$$\mathbf{H} = \mathbf{H}_M + \mathbf{H}_\mu + \mathbf{H}_{M\mu}. \quad (7)$$

Following methods used in the calculations for Compton scattering (Koba and Takeda 1948, Corinaldesi and Jost 1948, Schafroth 1949, Heitler and Ma 1949) we now perform a canonical transformation

$$\tilde{\mathbf{H}} = e^{-S} \mathbf{H} e^{+S}, \quad (8)$$

with $[\mathbf{H}_M + \mathbf{H}_\mu, S] + \mathbf{H}_{M\mu} = 0$. Using this relation we have

$$\tilde{\mathbf{H}} = \mathbf{H}_M + \mathbf{H}_\mu + \frac{1}{2}[\mathbf{H}_{M\mu}, S] + \frac{1}{3}[[\mathbf{H}_{M\mu}, S], S] + \frac{1}{8}[[[\mathbf{H}_{M\mu}, S], S], S] + \dots, \quad (9)$$

in which terms of first order have disappeared. The term $[\frac{1}{2}\mathbf{H}_{M\mu}, S]$ gives rise to the second-order matrix elements for scattering of a meson by a nucleon, which will be considered later, and to the self-energy of mesons and nucleons.

The term of the third order makes no contribution to the process under consideration. The term $\frac{1}{8}[[[\mathbf{H}_{M\mu}, S], S], S]$ gives a matrix element of fourth order for scattering of a meson by a nucleon, containing divergent integrals, and will be considered in detail in the following paragraphs.

It must be mentioned here that, if we use for the matrix elements of S the expression obtained from $[H_M + H_\mu, S] + H_{M\mu} = 0$ and put ourselves in the reference system in which $H_M + H_\mu$ is diagonal with eigenvalues E_p , i , e .

$$(k|S|n) = \frac{(k|H_{M\mu}|n)}{E_n - E_k},$$

the matrix element $(f|\frac{1}{8}[[[H_{M\mu}, S], S], S]|i)$ can be written

$$\begin{aligned} & (f|\frac{1}{8}[[[H_{M\mu}, S], S], S]|i) \\ &= \sum_{l,n} \sum_{m \neq i,f} \frac{(f|H_{M\mu}|l)(l|H_{M\mu}|m)(m|H_{M\mu}|n)(n|H_{M\mu}|i)}{(E_i - E_l)(E_i - E_m)(E_i - E_n)} \\ & \quad - \frac{1}{2} \sum_l \frac{(f|H_{M\mu}|l)(l|H_{M\mu}|i)}{E_i - E_l} \left\{ \sum_n \frac{(i|H_{M\mu}|n)(n|H_{M\mu}|i)}{(E_i - E_n)^2} \right. \\ & \quad \left. + \sum_n \frac{(f|H_{M\mu}|n)(n|H_{M\mu}|f)}{(E_f - E_n)^2} \right\} - \frac{1}{2} \sum_i \frac{(f|H_{M\mu}|l)(l|H_{M\mu}|i)}{(E_i - E_l)^2} \\ & \quad \times \left\{ \sum_n \frac{(i|H_{M\mu}|n)(n|H_{M\mu}|i)}{(E_i - E_n)} + \sum_n \frac{(f|H_{M\mu}|n)(n|H_{M\mu}|f)}{(E_f - E_n)} \right\}. \end{aligned} \quad (10)$$

Here i and f indicate the initial and final states, respectively, and l , m , n the intermediate states.

We have used the conservation of the energy between the initial and the final states, $E_i = E_f$, and the fact that $H_{M\mu}$ has vanishing matrix elements on the energy shell.

In what follows it will be necessary to know what would be the result of the S -transformation when applied to a term of second order in "g" in the initial Hamiltonian. If we suppose that this term has *vanishing matrix elements* for the *special process* treated in this paper, we should have, in analogy with (10):

$$\begin{aligned} & \sum_{l,m \neq i,f} \frac{1}{(E_i - E_l)(E_i - E_m)} (f|H_2|l)(l|H_{M\mu}|m)(m|H_{M\mu}|i) \\ & \quad + (f|H_{M\mu}|l)(l|H_2|m)(m|H_{M\mu}|i) + (f|H_{M\mu}|l)(l|H_{M\mu}|m)(m|H_2|i) \\ & \quad - \frac{1}{2} \sum_l \frac{(f|H_{M\mu}|l)(l|H_{M\mu}|i)}{E_i - E_l} \{ (i|H_2|i) + (f|H_2|f) \}. \end{aligned} \quad (11)$$

We shall call the terms of the first sums in (10) and (11) *a-terms*, and those of the other lines *b-terms*.

§ 2. SECOND-ORDER MATRIX ELEMENT FOR SCATTERING.

In this paper we shall study the process

$$(\mathbf{p}_i|+)(\mathbf{P}_i|P_+) \rightarrow (\mathbf{p}_f|+)(\mathbf{P}_f|P_+), \quad . \quad . \quad . \quad (12)$$

where $(\mathbf{p}|+)$ indicates a positive meson with momentum \mathbf{p} , $(\mathbf{p}|-)$ a negative meson with momentum \mathbf{p} , $(\mathbf{P}|P_+)$ a proton with momentum \mathbf{P} ,

$(\mathbf{P} | \mathbf{P}_-)$ a negative proton with momentum \mathbf{P} , $(\mathbf{P} | \mathbf{N}_+)$ a "positive" (*i. e.* ordinary) neutron with momentum \mathbf{P} and $(\mathbf{P} | \mathbf{N}_-)$ a "negative" neutron with momentum \mathbf{P} , *i. e.* a hole in the distribution of neutrons in the negative energy states. Occasionally we shall also use the notations \mathbf{P}_+ , \mathbf{P}_- , \mathbf{N}_+ , \mathbf{N}_- when there is no need to indicate the momentum.

Thus hole theory is used for the nucleons, and both meson field and proton-nucleon field are quantized.

The second-order matrix element for the process (12) is given by

$$-\frac{4\pi g^2}{2V\sqrt{(\omega_i\omega_f)}} \left\{ \sum_r \frac{(a^{*f}(\mathbf{P}_f)\beta\gamma_5 a^r(\bar{\mathbf{P}}))(a^{*r}(\bar{\mathbf{P}})\beta\gamma_5 a^i(\mathbf{P}_i))}{\Omega_i - \omega_f - \bar{\Omega}} \right. \\ \left. + \sum_r \frac{(a^{*f}(\mathbf{P}_f)\beta\gamma_5 b^r(\bar{\mathbf{P}}))(b^{*r}(\bar{\mathbf{P}})\beta\gamma_5 a^i(\mathbf{P}_i))}{-\omega_i + \Omega_f + \bar{\Omega}} \right\}, \quad . \quad . \quad . \quad (13)$$

where V is the periodicity volume, $\mathbf{P}_i - \mathbf{p}_f = \bar{\mathbf{P}}$, $\omega_i = \sqrt{(\mathbf{p}_i^2 + \mu^2)}$, $\omega_f = \sqrt{(\mathbf{p}_f^2 + \mu^2)}$, $\Omega_i = \sqrt{(\mathbf{P}_i^2 + M^2)}$, $\Omega_f = \sqrt{(\mathbf{P}_f^2 + M^2)}$, $\bar{\Omega} = \sqrt{(\bar{\mathbf{P}}^2 + M^2)}$ and $a(\mathbf{P})$, $b(\mathbf{P})$ satisfy the Dirac equations

$$\left. \begin{aligned} [(\boldsymbol{\alpha} \cdot \mathbf{P}) + \beta M] a(\mathbf{P}) &= \Omega(\mathbf{P}) a(\mathbf{P}), \\ [-(\boldsymbol{\alpha} \cdot \mathbf{P}) + \beta M] b(\mathbf{P}) &= -\Omega(\mathbf{P}) b(\mathbf{P}). \end{aligned} \right\} \quad . \quad . \quad . \quad (14)$$

(13) reduces to

$$(f | \mathbf{H}_2 | i) = a^{*f}(\mathbf{P}_f) \mathcal{H}_2 a^i(\mathbf{P}_i), \quad . \quad . \quad . \quad . \quad (15)$$

where

$$\mathcal{H}_2 = -\frac{2\pi g^2}{V\sqrt{(\omega_i\omega_f)}} \left\{ \frac{\omega_f - \Omega_i + \boldsymbol{\alpha} \cdot \bar{\mathbf{P}} + M\beta}{(\Omega_i - \omega_f)^2 - \bar{\Omega}^2} \right\}. \quad . \quad . \quad . \quad (16)$$

§ 3. SELF-ENERGIES.

The following self-energies of mesons and nucleons will occur :

(i) Self-energy of a meson with momentum \mathbf{p} .

Using perturbation theory and ordinary energy momentum representation we have

$$E^{\text{self}}(\mathbf{p}) = \frac{g^2}{2\pi\omega(\mathbf{p})} \int d\mathbf{P}' \frac{\Omega(\mathbf{P}') + \Omega(\mathbf{p} + \mathbf{P}')}{\Omega(\mathbf{P}')\Omega(\mathbf{p} + \mathbf{P}')} \frac{\Omega(\mathbf{P}')\Omega(\mathbf{p} + \mathbf{P}') + (\mathbf{P}' \cdot \mathbf{p} + \mathbf{P}') + M^2}{\omega^2(\mathbf{p}) - [\Omega(\mathbf{P}') + \Omega(\mathbf{p} + \mathbf{P}')]^2}. \quad . \quad . \quad . \quad (17)$$

This elementary treatment is along the lines used for Compton scattering which have been published up to date. No use is thus made of the new computational methods evolved by Dyson (1949). We introduce the Pauli-Rose coordinates (Pauli and Rose 1936)

$$\left. \begin{aligned} \mathbf{v} &= \frac{1}{2}[\Omega(\mathbf{p} + \mathbf{P}') - \Omega(\mathbf{P}')], \\ \mathbf{w} &= \frac{1}{2}[\Omega(\mathbf{p} + \mathbf{P}') + \Omega(\mathbf{P}')]. \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad (18)$$

After integration over v , this reduces to an integral over w which can be written in the form

$$E^{\text{self}}(\mathbf{p}) = \frac{8M^2g^2}{\pi} \cdot \frac{1}{\omega(\mathbf{p})} \int_1^\infty \frac{\sqrt{(\xi^2-1)\xi^2} d\xi}{\left(\frac{\mu}{M}\right)^2 - 4\xi^2}, \quad \dots \quad (19)$$

where
$$M^2\xi^2 = \left(w^2 - \frac{p^2}{4}\right). \quad \dots \quad (20)$$

We see from this result that the above *definition* of the improper integrals, which consists in integrating in the Pauli-Rose v - w -space, gives the correct covariance property for the self-energy.

A similar result has been obtained by S. T. Ma for the case of vacuum polarization in quantum electrodynamics (Ma 1949).

The integral (19) is independent of the momentum of the meson, and, calling $\delta\mu$ the self-energy of a meson at rest, we have

$$\mu\delta\mu = \omega(\mathbf{p})E^{\text{self}}(\mathbf{p}). \quad \dots \quad (21)$$

As far as this self-energy is concerned, it is superfluous for our purpose to introduce a process of regularization. The only important point is to have a definition of the self-energy with the correct covariance property, which is essential for the renormalization.

(ii) Self-energy of a nucleon with momentum \mathbf{P} .

Again, using ordinary perturbation theory, we obtain

$$E^{\text{self}}(\mathbf{P}) = (a^*(\mathbf{P})\beta a(\mathbf{P})) \frac{Mg^2}{4\pi} \left(\frac{1}{3} + N_\mu + \mu^2 X_\mu\right) - \frac{g^2}{12\pi} \Omega(\mathbf{P})(a^*(\mathbf{P})a(\mathbf{P})), \quad \dots \quad (22)$$

where N_μ and X_μ are functions given by Feynman (1948).

For the present purpose it is only necessary to state that N_μ contains a logarithmically divergent term, but both N_μ and X_μ are independent of the momentum \mathbf{P} .

The second term in the above self-energy has not the correct covariance property, and therefore a regularization method (Pauli and Villars 1949) must be introduced. Feynman's method is quite sufficient here. It consists in replacing the ordinary integration in the momentum space by a four-dimensional integration.

$$\int \dots \frac{d\mathbf{p}'}{\omega(\mathbf{p}')} = 2 \int \dots d\omega' d\mathbf{p}' \delta(\omega'^2 - \mathbf{p}'^2 - \mu^2), \quad \dots \quad (23)$$

putting
$$\delta(\omega'^2 - \mu_1^2 - \mathbf{p}'^2) \rightarrow \delta(\omega'^2 - \mu^2 - \mathbf{p}'^2) - \delta(\omega'^2 - \mu_1^2 - \mathbf{p}'^2), \quad \dots \quad (24)$$

and letting $\mu_1 \rightarrow \infty$ afterwards.

This type of regularization evidently eliminates the second term of our self-energy, leaving an expression of the form

$$E^{\text{self}}(\mathbf{P}) = A(a^*(\mathbf{P})\beta a(\mathbf{P})), \quad \dots \quad (25)$$

where A is independent of \mathbf{P} . It is not necessary here to give the expression for A explicitly. This process of regularization does not affect the finite matrix elements given in § 5.

§4. FOURTH-ORDER MATRIX ELEMENTS.

As has been mentioned above, the third-order matrix elements do not contribute to our process. Thus our cross-section will be proportional to

$$\begin{aligned} & |(f|H_2|i) + (f|H_4|i) + \dots|^2 \\ &= |(f|H_2|i)|^2 + (f|H_2|i)(f|H_4|i)^* + (f|H_2|i)^*(f|H_4|i) + \dots \\ &= H_{2,2} + H_{2,4} + H_{4,2} + \dots, \quad \dots \quad (26) \end{aligned}$$

where the meaning of $H_{2,2}$, $H_{2,4}$ is obvious.

The calculation of $(f|H_4|i)$ is extremely involved and we cannot give here full details and classify all the transitions.

It is possible, however, to give an account of the results in the following way:

(i) Transitions which give rise to integrals in the momentum space of the nucleons, while no mesons with variable momenta are present in any intermediate states. We found that in all the possible transitions of this type the mesons in the intermediate states are either identical with the initial and final meson, or have opposite charge and momentum. All the corresponding matrix elements are diverging. A first group of these integrals can be represented *exactly* as

$$\left(-\frac{\Gamma}{2\omega_i} E^{\text{self}}(\mathbf{p}_i) - \frac{1}{2\omega_f} E^{\text{self}}(\mathbf{p}_f) \right) (f|\frac{1}{2}[H_{M\mu}, S]|i), \quad \dots \quad (27)$$

where, by $E^{\text{self}}(\mathbf{p}_i)$ and $E^{\text{self}}(\mathbf{p}_f)$, we indicate the expressions obtained from (17) by substituting \mathbf{p}_i and \mathbf{p}_f for \mathbf{p} . Introducing for these integrals the same *definition* already used for the self-energies, *i. e.* expressing (17) by means of (19), we can write instead of (27) the expression

$$-\frac{1}{2}\mu\delta\mu \left(\frac{1}{\omega_i^2} + \frac{1}{\omega_f^2} \right) (f|\frac{1}{2}[H_{M\mu}, S]|i), \quad \dots \quad (28)$$

where $\delta\mu$ is the self-energy of the meson at rest. In order to subtract this term let us introduce into the Hamiltonian a term corresponding to a meson mass renormalization

$$H_{\delta\mu} = -2\mu\delta\mu \int \psi^* \psi \, d\tau. \quad \dots \quad (29)$$

This new $H_{\delta\mu}$ gives rise to the fourth-order transitions

$$\begin{aligned} & i \xrightarrow{(H_{M\mu})} (\mathbf{p}_i|+) (\mathbf{p}_f|+) (\bar{\mathbf{P}}|N_+) \xrightarrow{(H_{M\mu})} f(\mathbf{p}_i|+) (-\mathbf{p}_i|-) \xrightarrow{(H_{\delta\mu})} f, \\ & i \xrightarrow{(H_{\delta\mu})} i(\mathbf{p}_f|+) (-\mathbf{p}_f|-) \xrightarrow{(H_{M\mu})} (\mathbf{p}_i|+) (\mathbf{p}_f|+) (\bar{\mathbf{P}}|N_+) \xrightarrow{(H_{M\mu})} f. \end{aligned} \quad \dots \quad (30)$$

(30) corresponds to the second-order transition for $(f|\frac{1}{2}[H_{M\mu}, S]|i)$

$$i \rightarrow (\mathbf{p}_i|+) (\mathbf{p}_f|+) (\bar{\mathbf{P}}|N_+) \rightarrow f, \quad \dots \quad (31)$$

(giving rise to the first term in (13)), and

$$\left. \begin{aligned}
 i \rightarrow (H_{M\mu}) \rightarrow (\mathbf{p}_i|+)(-\mathbf{p}_i|-)(\mathbf{P}_i|P_+)(\mathbf{P}_f|P_+)(-\bar{\mathbf{P}}|N_-) \rightarrow (H_{M\mu}) \rightarrow \\
 \rightarrow f(\mathbf{p}_i|+)(-\mathbf{p}_i|-) \rightarrow (H_{\delta\mu}) \rightarrow f, \\
 i \rightarrow (H_{M\mu}) \rightarrow (\mathbf{p}_i|+)(-\mathbf{p}_i|-)(\mathbf{P}_i|P_+)(\mathbf{P}_f|P_+)(-\bar{\mathbf{P}}|N_-) \rightarrow (H_{\delta\mu}) \rightarrow \\
 \rightarrow (\mathbf{P}_i|P_+)(\mathbf{P}_f|P_+)(-\bar{\mathbf{P}}|N_-) \rightarrow (H_{M\mu}) \rightarrow f, \\
 i \rightarrow (H_{\delta\mu}) \rightarrow i(\mathbf{p}_f|+)(-\mathbf{p}_f|-) \rightarrow (H_{M\mu}) \rightarrow \\
 \rightarrow (\mathbf{p}_f|+)(-\mathbf{p}_f|-)(\mathbf{P}_i|P_+)(\mathbf{P}_f|P_+)(-\bar{\mathbf{P}}|N_-) \rightarrow (H_{M\mu}) \rightarrow f, \\
 i \rightarrow (H_{M\mu}) \rightarrow (\mathbf{P}_i|P_+)(\mathbf{P}_f|P_+)(-\bar{\mathbf{P}}|N_-) \rightarrow (H_{\delta\mu}) \rightarrow \\
 \rightarrow (\mathbf{p}_f|+)(-\mathbf{p}_f|-)(\mathbf{P}_i|P_+)(-\bar{\mathbf{P}}|N_-) \rightarrow (H_{M\mu}) \rightarrow f.
 \end{aligned} \right\} \quad \dots (32)$$

(32) corresponds to the second-order transition

$$i \rightarrow (\mathbf{P}_i|P_+)(\mathbf{P}_f|P_+)(\mathbf{P}|N_+) \rightarrow f, \quad \dots (33)$$

(giving rise to the second term in (13)). The contribution due to the corresponding matrix elements gives exactly the expression (28) with opposite sign.

Another group of transitions gives an expression which can be written

$$\begin{aligned}
 & -\frac{\pi g^2}{V \sqrt{(\omega_i \omega_f)}} \left\{ \sum_r \frac{(\dots)(\dots)}{[\Omega_i - \omega_f - \bar{\Omega}]^2} \right\} \frac{(\omega_i + \omega_f) \mu \delta\mu}{\omega_i \omega_f} \quad \dots (34) \\
 & \left(\text{from } E^{\text{self}}(\mathbf{p}_i) + E^{\text{self}}(\mathbf{p}_f) = \frac{\omega_i + \omega_f}{\omega_i \omega_f} \mu \delta\mu \right),
 \end{aligned}$$

where the brackets $(\dots)(\dots)$ indicate the various factors in the numerators of the first term in (13). This expression is subtracted exactly by the transition

$$i \rightarrow (\mathbf{p}_i|+)(\mathbf{p}_f|+)(\bar{\mathbf{P}}|N_+) \Rightarrow (\mathbf{p}_i|+)(\mathbf{p}_f|+)(\bar{\mathbf{P}}|N_+) \rightarrow f, \quad \dots (35)$$

where \rightarrow indicates the diagonal matrix elements, *i.e.* the expectation value of $-2\mu \delta\mu \int \psi^* \psi d\tau$ for a state consisting of vacuum plus $(\mathbf{p}_i|+)$ (or, respectively, of vacuum plus $(\mathbf{p}_f|+)$). These are the self-energies of mesons in the intermediate state

$$(\mathbf{p}_i|+)(\mathbf{p}_f|+)(\bar{\mathbf{P}}|N_+)$$

of the second-order process (in the other intermediate state no mesons are present).

Finally, we have had to consider carefully contributions due to b-terms of (10) (third term) and (11), which also go away exactly.

With this procedure we are able to subtract completely all the matrix elements of the type "mass corrections for mesons". It is also possible

to verify that *all* the above-mentioned matrix elements, which are cancelled out by various types of transitions due to the counter-term (29), can be written as

$$\frac{\partial}{\partial \omega_i} (f | \frac{1}{2} [\mathbf{H}_{M\mu}, S] | i) E^{\text{self}}(\mathbf{p}_i) + \frac{\partial}{\partial \omega_f} (f | \frac{1}{2} [\mathbf{H}_{M\mu}, S] | i) E^{\text{self}}(\mathbf{p}_f) \quad (36)$$

which gives

$$\frac{\partial}{\partial \mu} (f | \frac{1}{2} [\mathbf{H}_{M\mu}, S] | i) \cdot \delta \mu \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

if we use the covariance property of the self-energy (19).

(ii) We must now consider transitions which give rise to matrix elements containing integrals in the momentum space of mesons. (By this we mean that, in at least one of the intermediate states, there is a meson with variable momentum.)

First we find a group of matrix elements all containing as a common factor, either

$$(a^*(\mathbf{P}_f) \dots b(-\mathbf{P}_f)), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38a)$$

or

$$(b^*(-\mathbf{P}_i) \dots a(\mathbf{P}_i)). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38b)$$

The terms (38a) are identical with what is obtained in a formal way by replacing $a^*(\mathbf{P})$ by $a^*(\mathbf{P}_f)$ and $a(\mathbf{P})$ by $b(-\mathbf{P}_f)$, in those integrals which represent the self-energy of a proton, *before* evaluating them. (38b) has a similar meaning.

If for (38a, b) we carry out the integration in the *same manner* as we did in order to get (22), and if we follow the *same* rule for the regularization, we obtain

$$\{(b^*(-\mathbf{P}_i)\beta a(\mathbf{P}_i)) \text{ or } (a^*(\mathbf{P}_f)\beta b(-\mathbf{P}_f))\} R. \quad . \quad . \quad . \quad (39)$$

R stands for the result of regularizing the expression

$$\frac{M^2 g^2}{4\pi} (\frac{1}{3} + N_\mu + \mu^2 X_\mu).$$

The second term in (22) has no term corresponding to it in (39) because of the orthogonality of $a^*(\mathbf{P})$ and $b(\mathbf{P})$. However, since it is eliminated by regularization, the similarity between (22) and (39) is re-established. We can verify that the group of matrix elements we are treating now is just subtracted *exactly* if we introduce into the Hamiltonian the counter term

$$H_{\delta M} = -\delta M \int \Phi_P^* \beta \Phi_P d\tau. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

Here δM is the self-energy of a proton at rest and Φ_P is the above mentioned operator describing the protons. We consider the following

transitions due to (40):

$$\left. \begin{aligned} i \rightarrow (H_{M\mu}) \rightarrow i(\mathbf{p}_f | +)(-\mathbf{p}_i | P_-)(\bar{\mathbf{P}} | N_+) \rightarrow (H_{M\mu}) \rightarrow \\ \rightarrow f(\mathbf{p}_i | P_+)(-\mathbf{p}_i | P_-) \rightarrow (H_{\delta M}) \rightarrow f, \\ i \rightarrow (H_{M\mu}) \rightarrow i(\mathbf{p}_f | +)(-\mathbf{p}_i | P_-)(\bar{\mathbf{P}} | N_+) \rightarrow (H_{\delta M}) \rightarrow \\ \rightarrow (\mathbf{p}_i | +)(\mathbf{p}_f | +)(\bar{\mathbf{P}} | N_+) \rightarrow (H_{M\mu}) \rightarrow f, \\ i \rightarrow (H_{M\mu}) \rightarrow (\mathbf{p}_i | P_+)(\mathbf{p}_f | P_+)(-\bar{\mathbf{P}} | N_-) \rightarrow (H_{M\mu}) \rightarrow \\ \rightarrow f(\mathbf{p}_i | P_+)(-\mathbf{p}_i | P_-) \rightarrow (H_{\delta M}) \rightarrow f \end{aligned} \right\} \quad (41)$$

and

$$\left. \begin{aligned} i \rightarrow (H_{\delta M}) \rightarrow i(\mathbf{p}_f | P_+)(-\mathbf{p}_f | P_-) \rightarrow (H_{M\mu}) \rightarrow \\ \rightarrow f(\mathbf{p}_i | +)(-\mathbf{p}_f | P_-)(\bar{\mathbf{P}} | N_+) \rightarrow (H_{M\mu}) \rightarrow f, \\ i \rightarrow (H_{M\mu}) \rightarrow (\mathbf{p}_i | +)(\mathbf{p}_f | +)(\bar{\mathbf{P}} | N_+) \rightarrow (H_{\delta M}) \rightarrow \\ \rightarrow f(\mathbf{p}_i | +)(-\mathbf{p}_f | P_-)(\bar{\mathbf{P}} | N_+) \rightarrow (H_{M\mu}) \rightarrow f, \\ i \rightarrow (H_{\delta M}) \rightarrow i(\mathbf{p}_f | P_+)(-\mathbf{p}_f | P_-) \rightarrow (H_{M\mu}) \rightarrow \\ \rightarrow (\mathbf{p}_i | P_+)(\mathbf{p}_f | P_+)(-\bar{\mathbf{P}} | N_-) \rightarrow (H_{M\mu}) \rightarrow f. \end{aligned} \right\} \quad (42)$$

Some other matrix elements cancel out *exactly* with the transitions, due to (40):

$$\begin{aligned} i \rightarrow (H_{M\mu}) \rightarrow (\mathbf{p}_i | P_+)(\mathbf{p}_f | P_+)(-\bar{\mathbf{P}} | N_-) \Rightarrow \\ \Rightarrow (\mathbf{p}_i | P_+)(\mathbf{p}_f | \bar{P}_+)(-\mathbf{P} | N_-) \rightarrow (H_{M\mu}) \rightarrow f, \quad (43) \end{aligned}$$

where \Rightarrow indicates the diagonal matrix elements of (40) either for a state consisting of vacuum plus a proton $(\mathbf{p}_i | P_+)$, or for a state consisting of vacuum plus a proton $(\mathbf{p}_f | P_+)$.

Other matrix elements are cancelled out exactly if we consider carefully the contributions due to the third term in (10) and the second term in (11). We consider it unnecessary to enumerate all these terms in detail.

Finally, there remain divergent matrix elements both in the momentum space of nucleons and of mesons. They can be written in the form

$$(f | \frac{1}{2}[H_{M\mu}, S] | i) \frac{2(\delta g_1 + \delta g_2)}{g} \quad (44)$$

δg_1 is a logarithmically divergent integral in the momentum space of nucleons, *i. e.* it is due to the transitions which do not involve mesons with variable momenta. δg_1 can be evaluated in the same way as (19) and reduces to

$$\frac{2}{g} \delta g_1 = - \frac{16g^2}{\pi} \int_1^\infty \frac{\sqrt{(\xi^2 - 1)} \xi^2 d\xi}{\left(4\xi^2 - \frac{\mu^2}{M^2}\right)^2} \quad (45)$$

δg_2 is given by the expression

$$\frac{2}{g} \delta g_2 = \frac{g^2}{4\pi} \left[N_\mu + (\mu^2 + M^2) X_\mu(M, M) + \frac{M\mu^2}{2} \left(\frac{\partial}{\partial M_0} X_\mu(M, M_0) \right)_{M_0=M} \right]. \quad (46)$$

It is due to contributions of the second term in (10). The notations are as in Feynman's paper (Feynman 1948). The expression for δg_2 is obtained by evaluating integrals in the momentum space of mesons in the same manner as for the self-energy of protons. δg_1 and δg_2 are independent of the momenta of the particles taking part in our process. The whole expression (44) can be subtracted completely by adding to the Hamiltonian a term $-(\delta g_1 + \delta g_2)(H_{M\mu}/g)$. We can regard $\delta g_1 + \delta g_2$ as a renormalization (or rather part of it) of the mesonic charge g .

(iii) In (i) and (ii) we have frequently used the word *exactly* in order to indicate that certain matrix elements of the fourth-order could be subtracted *completely* as mass corrections or mesonic charge corrections. There are now further divergent contributions met with in this investigation, of which only the divergent parts can be subtracted, but in which some finite parts are left over. These divergences are connected with the self-energy of the *neutron* in the intermediate state, and a further contribution to δg . We introduce into the Hamiltonian a term similar to (40), referring *neutrons*:

$$H_{\delta M} = -\delta M \int \Phi_N^* \beta \Phi_N d\tau, \quad . \quad . \quad . \quad . \quad (40')$$

(where δM is the self-energy of a neutron at rest and Φ_N^* , Φ_N are the operators describing the neutrons) and a term

$$-\delta g'_2 \frac{H_{M\mu}}{g},$$

representing a further contribution to the renormalization of g . It turns out that $\delta g'_2 = \delta g_2$, given by (46).

The total correction to g is thus

$$\delta g = \delta g_1 + \delta g_2 + \delta g'_2 = \delta g_1 + 2\delta g_2.$$

The counter-term (40') gives rise to the following transitions:

(i) Transitions involving the absorption (or emission) of a pair of neutrons $(\bar{\mathbf{P}} | N_+)(-\bar{\mathbf{P}} | N_-)$. The divergences of these contributions cancel out with a considerable number of divergent matrix elements of the fourth-order, but some finite terms are left over. The evaluation of these finite terms in the extreme non-relativistic limit shows that they can be neglected compared with the other terms contained in (50) (§ 5).

(ii) Transitions involving the expectation value of the counter-term itself, for the state $\{\text{vacuum} + (\bar{\mathbf{P}} | N_+)\}$ or $\{\text{vacuum} + (-\bar{\mathbf{P}} | N_-)\}$. By combining these transitions with the contributions due to the term involving $\delta g'_2$ mentioned above, we can eliminate all the remaining divergences, but a finite part is left, which is included in (50) (§ 5).

We are indebted to Professor Heitler for pointing out to us that for the consistency of the above interpretation of δg it would be necessary to carry out the calculation of the vacuum polarization due to free nucleons in a way similar to that of deriving the electronic charge renormalization in electrodynamics. This should lead to the same value of δg . This has not been done here. We have restricted ourselves to the proof that part of the divergences can be expressed in the form δg , where δg is constant and invariant.

§ 5. FOURTH-ORDER, FINITE, MATRIX ELEMENTS.

In addition to the finite parts left in the fourth-order transitions considered in § 4 (iii), there is a group of some thirty others which give rise to finite matrix elements, and thus contribute to the higher-order terms of the scattering cross-section. In all these transitions, mesons with variable momentum \mathbf{p}' are present in the intermediate states. Only the extreme non-relativistic case will be discussed here, for which

$$\mathbf{P}_i, \mathbf{p}_i, \mathbf{P}_f, \mathbf{p}_f \ll \mu. \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

This over-simplified, but still involved, case is given in anticipation of a more refined non-relativistic treatment, which will be discussed in Part II. of this paper. The assumptions (47) enable us to obtain only a low-energy *limit* of the cross-section.

For the case (47), only eleven transitions remain which give non-vanishing contributions of the fourth-order. They all involve the virtual emission and re-absorption of three nucleons in the following way :

$$\begin{aligned} &\text{either : one } P_+ \text{ and two } N_- ; \\ &\text{or} \quad : \text{one } P_- \text{ and two } N_+. \end{aligned}$$

It is seen from (26) that the terms of lowest order containing H_4 are

$$(f | H_2 | i)(f | H_4 | i)^* + (f | H_2 | i)^*(f | H_4 | i) = H_{2,4} + H_{4,2}. \quad (48)$$

Summing over the final, and averaging over the initial, spin states respectively, (48) reduces to

$$H_{2,4} + H_{4,2} = \frac{1}{4} \text{Sp} \{ \mathcal{H}_2 (1 + \beta) \mathcal{H}_4 (1 + \beta) \}, \quad . \quad . \quad . \quad (49)$$

for the particular case (47). Here, \mathcal{H}_4 is given by

$$\begin{aligned} \mathcal{H}_4 = & \frac{\pi^2 g^4}{2V^2 \mu} \\ & \times \sum_{\mathbf{p}'} \frac{\beta \gamma_5 [\Omega(\mathbf{p}_1) \mp \boldsymbol{\alpha} \cdot \mathbf{p}_1 \pm M\beta] \beta \gamma_5 [\Omega(\mathbf{p}_2) \pm \boldsymbol{\alpha} \cdot \mathbf{p}_2 \mp M\beta] \beta \gamma_5 [\Omega(\mathbf{p}_3) \mp \boldsymbol{\alpha} \cdot \mathbf{p}_3 \pm M\beta]}{\omega(\mathbf{p}') \Omega(\mathbf{p}_1) \Omega(\mathbf{p}_2) \Omega(\mathbf{p}_3) (E_i - E_l)(E_i - E_m)(E_i - E_n)}, \end{aligned} \quad (50)$$

and \mathcal{H}_2 is obtained from (16) by inserting the values (47). In formula (50),

$$\begin{aligned} \mathbf{p}_2 &= \mathbf{p}' \\ \mathbf{p}_1 &= \mathbf{p}_3 = \mathbf{p}' \quad \text{or} \quad 0. \end{aligned}$$

Integrating over momentum space and combining the results for all transitions (including those mentioned in 4(iii))[†],

$$H_{2,4} + H_{4,2} + \dots - \frac{3\pi g^6}{4V^2\mu^2 M^2} \left(2 + \frac{3\mu}{M}\right). \quad \dots \quad (52)$$

$|H_2|^2$ can be readily evaluated from (16). Summing, as before, over the final, and averaging over the initial, spins,

$$|H_2|^2 = \frac{\pi^2 g^4}{V^2\mu^2 M^2} \left(1 + \frac{\mu}{M}\right). \quad \dots \quad (53)$$

Now, the scattering probability per unit time, w , is defined by

$$w = 2\pi\rho(\omega) |H|^2, \quad \dots \quad (54)$$

where

$$\rho(\omega) = \frac{1}{(2\pi)^3} \omega p V \sin\theta \, d\theta \, d\phi$$

(where the element of solid angle is $\sin\theta \, d\theta \, d\phi$).

The differential cross-section becomes in our case

$$\begin{aligned} d\sigma &= d\sigma_4 + d\sigma_6 + \dots \\ &= \frac{\sin\theta \, d\theta \, d\phi}{(2\pi)^2} \mu^2 V^2 (|H_2|^2 + H_{2,4} + H_{4,2} + \dots). \quad \dots \quad (55) \end{aligned}$$

Thus the fourth-order term is

$$d\sigma_4 = \frac{1}{4M^2} g^4 \sin\theta \, d\theta \, d\phi \left(1 + \frac{\mu}{M}\right). \quad \dots \quad (56)$$

In the limit $p \ll \mu$, the cross-section therefore tends to a *finite* value. The sixth-order term is

$$d\sigma_6 = -\frac{3}{16\pi M^2} g^6 \sin\theta \, d\theta \, d\phi \left(2 + \frac{3\mu}{M}\right), \quad \dots \quad (57)$$

which is also independent of p .

Therefore the ratio of the sixth- and fourth-order terms in the expression for the cross-section is given by

$$\left| \frac{d\sigma_6}{d\sigma_4} \right| = \frac{3}{4\pi} g^2 \left(2 + \frac{\mu}{M}\right). \quad \dots \quad (58)$$

For $\mu/M = \frac{1}{6}$ this becomes

$$\left| \frac{d\sigma_6}{d\sigma_4} \right| = \frac{g^2}{1.9}.$$

In order to obtain a numerical value for $|d\sigma_6/d\sigma_4|$, and thus a measure of the convergence of the expansion in terms of g^2 , a knowledge

[†] The fact that $H_{2,4} + H_{4,2} < 0$, and therefore $d\sigma_6 < 0$, does not lead to a self-contradictory result, so long as $d\sigma > 0$. We have verified that $|H_4|^2$ is of such magnitude that $|H_2 + H_4|^2 > 0$.

of g is required. For conventional values of g it seems that the second term in the series (55) is not very much smaller than the first term†. The case of longitudinal mesons has already been treated by purely non-relativistic calculations by Heitler and Ma (1949).

Note added in proof.—Dancoff and Drell (*Phys. Rev.* **76**, (2), 1949, 205–212) mention that the value of $g^2/4$ was suggested by Bethe at the Spring 1949 meetings of the American Physical Society in Washington, D.C. (B.A.P.S., vol. 24, No. 4).

§ 6. CONCLUSION.

The present investigation was carried out with a double purpose: First, to decide whether the subtraction methods, which have been successful in quantum electrodynamics, also work in meson theory. This we have shown to be the case for the pseudoscalar coupling considered (but it is doubtful whether the same is true for other types of coupling). Secondly, to check whether the method of expanding in powers of g can be applied to meson problems. This, in view of the slow convergence of the terms in question, seems rather doubtful, at any rate if results more accurate than the rough order of magnitude are wanted.

An extension of formula (52) for the higher energy region $p \sim \mu$ but $p \ll M$ will be given in Part II., together with a discussion of the damping effects.

The authors wish to express their gratitude to Professor W. Heitler, who suggested the problem, for his invaluable criticism, and to Dr. S. T. Ma, for his interest in the first portion of this paper.

REFERENCES.

- CORINALDESI, E., and JOST, R., 1948, *Helv. Phys. Acta*, **21**, 183.
 DYSON, F. J., 1949, *Phys. Rev.*, **75**, 486.
 FEYNMAN, R. P., 1948, *Phys. Rev.*, **74**, 1430.
 HEITLER, W., and MA, S. T., 1949, *Phil. Mag.*, **40**, 651.
 Koba, Z., and TAKEDA, G., 1948, *Progr. Theor. Phys.*, **3**, 98.
 MA, S. T., 1949, *Phys. Rev.*, **75**, 1265, and a detailed paper to be published.
 PAULI, W., and ROSE, M. E., 1936, *Phys. Rev.*, **49**, 462.
 PAULI, W., and VILLARS, F., 1949, *Rev. Mod. Phys.*, in the press.
 SCHAFROTH, R., 1949, *Phys. Rev.*, **75**, 1111.
 SLOTNICK, M., and HEITLER, W., 1949, *Phys. Rev.*, **75**, 1645.
 VILLARS, F., 1947, *Helv. Phys. Acta*, **20**, 476.
 WENTZEL, G., 1949, *Quantum Theory of Fields* (New York: Interscience Publishers, Inc.).

† Other values of g are, however, met with in current literature which would make the series diverge (Villars 1947, Slotnick and Heitler).

CVII. *Notes on the Validity and Application of the Method of
Molecular Orbitals.*

By C. A. COULSON* and H. C. LONGUET-HIGGINS†‡.

[Received August 10, 1949.]

§ 1.

THE method of molecular orbitals (m.o.) has been extensively used for the study of conjugated and aromatic molecules, both with and without the presence of hetero-atoms such as nitrogen and oxygen. Even in the simplest form introduced by Hückel, it has shown itself able to give a surprisingly good account of the ground states of these molecules; that is to say, it correctly describes their resonance, or delocalization, energies, and in addition, is able to predict bond lengths and charge distributions. Thus quite minute disparities in the observed bond lengths of coronene and pyrene find a satisfactory description in this theory (Moffitt and Coulson 1948); so do the charge migrations which govern the reactivity of heteromolecules of this type towards electrophilic or nucleophilic reagents (Coulson and Longuet-Higgins 1947 b). It seems as if, despite its extreme simplicity, the method is able to account fairly well for the ground state: yet it fails quite seriously with excited states, so that, for example, in naphthalene, it grossly misinterprets the nature of the three long-wave u.v. absorption bands (Miss Jacobs 1949). It is true that some of the factors responsible for this—that the m.o.'s used are probably not the best that could be imagined (Coulson and Miss Fischer 1949), and that there is a degree of configurational interaction not allowed for in the simple theory (Coulson, Craig and Miss Jacobs 1949)—operate both in the ground and excited states. But the agreement with observation and experiment shows plainly that in the ground state the method possesses a surprising validity. It is the purpose of our first paragraph to provide a partial explanation of this fact.

According to the method of m.o., each π electron occupies an orbital compounded linearly out of the appropriate atomic orbitals ϕ_r ($r=1, 2, \dots n$) of the constituent atoms 1, $\dots n$. Thus the j th m.o. would be written

$$\psi_j = \sum_{r=1}^n c_{jr} \phi_r, \quad \dots \dots \dots (1)$$

* Wheatstone Physics Laboratory, King's College, London.

† Chemistry Department, The University, Manchester.

‡ Communicated by the Authors.

It is quite straightforward to show that

$$\left. \begin{aligned} C_{jr}^2 + C_{kr}^2 &= c_{jr}^2 + c_{kr}^2, \\ C_{jr}C_{js} + C_{kr}C_{ks} &= c_{jr}c_{js} + c_{kr}c_{ks}. \end{aligned} \right\} \dots \dots \dots (8)$$

This shows that in (5) neither q_r nor p_{rs} is affected by the replacement of ψ_j and ψ_k by Ψ_j and Ψ_k . This process can be continued stage by stage until any general linear orthogonal transformation has been obtained. Indeed the general proof may be given in one step, but the simple argument above seemed more instructive. This completes the proof that neither q_r , p_{rs} nor E are affected by such operations.

This argument serves to reinforce what has already been said several times by other people about excited states. The failure here is essentially greater than for the ground state because of the inability of any simple product wave function (or any single determinantal wave function) to deal satisfactorily with spin degeneracy.

§ 2.

The result proved in §1 has an immediate application in cases where two or more of the m.o. are degenerate. For if ψ_j and ψ_k are degenerate, any linear combinations of them will also serve as allowed m.o. Our theorem shows us that provided all the degenerate m.o. are occupied by the same number of electrons (which must necessarily be either 1 or 2) then it makes no difference to our calculated bond orders and charges whether we use ψ_j and ψ_k or any orthogonal combination of them. This result is true, however great the degeneracy, but of course it is not true if the occupation numbers of the orbitals are not all equal.

§ 3.

Let us suppose that there is a set of m degenerate orthogonal orbitals ψ_f , each singly occupied. Then if we denote by $p_{rs}^{(f)}$ and $q_r^{(f)}$ the contributions of these orbitals to p_{rs} and q_r , it follows that

$$p_{rs}^{(f)} = \frac{1}{2} \sum_f (c_{fr}^* c_{fs} + c_{fs}^* c_{fr}), \quad \dots \dots \dots (9)$$

$$q_r^{(f)} = \sum_f c_{fr}^* c_{fr}. \quad \dots \dots \dots (10)$$

c^* is the complex conjugate of c , no longer necessarily supposed to be real.

Now denote by ψ_g all the m.o. except those of the subset ψ_f . This means that ψ_f and ψ_g together represent the complete molecular shell associated with the original given atomic orbitals. We could say (Coulson and Rushbrooke 1940) that together they form a complete set of orthonormal vectors in the space defined by the atomic orbitals ϕ . Thus orthogonality and normalization conditions show us that

$$\sum_f c_{fr}^* c_{fr} + \sum_g c_{gr}^* c_{gr} = 1, \quad \dots \dots \dots (11)$$

$$\sum_f c_{fr}^* c_{fr} + \sum_g c_{gr}^* c_{gs} = 0. \quad \dots \dots \dots (12)$$

Thus, from (9)–(12), we see that

$$p_{rs}^{(f)} = -\frac{1}{2} \sum_g (c_{gr}^* c_{gs} + c_{gs}^* c_{gr}), \quad . \quad . \quad . \quad . \quad . \quad (13)$$

$$q_r^{(f)} = 1 - \sum_g c_{gr}^* c_{gr}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

This shows that the contributions to p_{rs} and q_r from the electrons in the degenerate orbitals ψ_f may be expressed in terms of the coefficients c_{gr} of the remaining orbitals, always provided that ψ_f and ψ_g together form a complete set.

This gives us an alternative proof of the result in § 2. For any unitary transformation of the ψ_f will still ensure that together with the ψ_g we have a complete set. But (13) and (14) are independent of which particular linear combinations of the ψ_f we may choose. Such choice therefore has no effect upon p_{rs} and q_r . This was the theorem.

§ 4.

An alternative proof of (13) and (14), not involving explicit use of the orthogonality relations (11) and (12), may be found from a very simple result, which does not seem to have been stated in the literature, although it is implicit both in Coulson and Rushbrooke (1940) and in Coulson and Longuet-Higgins (1947 a). This result is that if there is one electron in each of the orbitals of a complete molecular shell, the resulting bond orders are all zero and the charges are all unity. The proofs of these are immediate if we use the notation of Coulson and Longuet-Higgins (particularly their equations 31–44) and recognize that

$$\int \frac{\Delta_{r,s}(z)}{\Delta(z)} dz, \quad \text{and} \quad \int \frac{\Delta_{r,r}(z)}{\Delta(z)} dz$$

round the complete circle at infinity are respectively equal to 0 and $-2\pi i$. If now we write (13) and (14) in the form

$$p_{rs}^{(f)} = -p_{rs}^{(g)}, \quad q_r^{(f)} = 1 - q_r^{(g)}, \quad . \quad . \quad . \quad . \quad . \quad (15)$$

it is seen that they are simply restatements of the two results referred to above.

§ 5.

Our last paragraph is concerned with an alternative way of calculating charges and bond orders when certain m.o. are degenerate. Let us suppose that in a given molecule there are m degenerate orthonormal m.o. ψ_f , each singly occupied, but that our calculations have only provided us with a set of m linearly independent m.o. ω_n , which are not mutually orthogonal, but which all do have the correct energy. In order to find p_{rs} and q_r we could of course employ Schmidt's orthogonalization process to construct from the ω_n an orthonormal set ψ_f , and then use (9) and (10). However, this is sometimes a rather lengthy procedure, and can be short-circuited in the following manner.

If we put

$$\omega_h = \sum_m a_{mh} \phi_m, \quad \psi_f = \sum_h b_{fh} \omega_h = \sum_t c_{ft} \phi_t,$$

then

$$c_{ft} = \sum_h b_{fh} a_{ht}.$$

Also, from the orthogonality of the ψ_f ,

$$\sum_t c_{ft} c_{f't}^* = \delta_{ff'}.$$

In matrix notation, if $(a_{ht}) = \mathbf{A}$, $(b_{fh}) = \mathbf{B}$, $(c_{ft}) = \mathbf{C}$, we have

$$\mathbf{C} = \mathbf{B}\mathbf{A}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

and

$$\mathbf{C}\mathbf{C}^\dagger = \mathbf{1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where \mathbf{C}^\dagger is the adjoint of \mathbf{C} , *i. e.* the transpose of its complex conjugate. Now by comparing (16) and (17) with (9) and (10) we see that the contributions of the electrons in ψ_f to q_r is simply the r th diagonal element of $\mathbf{C}^\dagger \mathbf{C}$; and their contribution to p_{rs} is the mean of the appropriate off-diagonal elements in the same matrix. Since the matrix \mathbf{B} is supposed unknown, our object is to determine the elements of $\mathbf{C}^\dagger \mathbf{C}$ in terms of the elements of \mathbf{A} , which are known. Now from (16) and (17)

$$\mathbf{B}\mathbf{A}\mathbf{A}^\dagger \mathbf{B}^\dagger = \mathbf{C}\mathbf{C}^\dagger = \mathbf{1}.$$

Multiplying on the left by \mathbf{B}^\dagger and on the right by \mathbf{B} ,

$$\mathbf{B}^\dagger \mathbf{B} \mathbf{A} \mathbf{A}^\dagger \mathbf{B}^\dagger \mathbf{B} = \mathbf{B}^\dagger \mathbf{B}.$$

If we divide by $\mathbf{B}^\dagger \mathbf{B}$ it follows that

$$\mathbf{B}^\dagger \mathbf{B} = (\mathbf{A}\mathbf{A}^\dagger)^{-1}.$$

Thus

$$\mathbf{C}^\dagger \mathbf{C} = \mathbf{A}^\dagger \mathbf{B}^\dagger \mathbf{B} \mathbf{A} = \mathbf{A}^\dagger (\mathbf{A}\mathbf{A}^\dagger)^{-1} \mathbf{A}. \quad . \quad . \quad . \quad . \quad . \quad (18)$$

This shows that the matrix $\mathbf{C}^\dagger \mathbf{C}$ whose elements give the electron densities and bond orders due to the electrons in ψ_f , are expressible directly in terms of the coefficients a_{ht} .

This takes a particularly simple form in the important case that the m.o. ψ_f are a doubly-degenerate pair, and that we can find two linearly independent combinations of them, *viz.*,

$$\omega_1 = \sum_t a_{1t} \phi_t, \quad \omega_2 = \sum_t a_{2t} \phi_t.$$

Then

$$\mathbf{A}\mathbf{A}^\dagger = \begin{pmatrix} \sum_t a_{1t} a_{1t}^* & \sum_t a_{1t} a_{2t}^* \\ \sum_t a_{2t} a_{1t}^* & \sum_t a_{2t} a_{2t}^* \end{pmatrix},$$

and after a little reduction, and the use of (18), it follows that the element in the r th row and the s th column of $\mathbf{C}^\dagger \mathbf{C}$ is

$$\frac{a_{1r}^* a_{1s} \sum_t a_{2t} a_{2t}^* - a_{1r}^* a_{2s} \sum_t a_{1t} a_{2t}^* - a_{2r}^* a_{1s} \sum_t a_{2t} a_{1t}^* + a_{2r}^* a_{2s} \sum_t a_{1t} a_{1t}^*}{(\sum_t a_{1t} a_{1t}^*)(\sum_t a_{2t} a_{2t}^*) - (\sum_t a_{1t} a_{2t}^*)(\sum_t a_{2t} a_{1t}^*)}. \quad . \quad . \quad (19)$$

Let us call this expression $f(r, s)$. Then

$$q_r^{(f)} = f(r, r), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

$$p_{rs}^{(f)} = \frac{1}{2} \{ f(r, s) + f(s, r) \}. \quad (21)$$

As a check on these equations it may be noted that if ω_1 and ω_2 happen to be orthogonal, then $\Sigma a_{1t} a_{2t}^* = \Sigma a_{2t} a_{1t}^* = 0$, so that (20) and (21) reduce at once to the usual definitions (9) and (10).

REFERENCES.

- COULSON, C. A., CRAIG, D. P. and Miss JACOBS (in press).
COULSON, C. A. and Miss FISCHER, 1949, *Phil. Mag.*, **40**, 386.
COULSON, C. A. and LONGUET-HIGGINS, H. C., 1947 a, *Proc. Roy. Soc. A*, **191**, 39; 1947 b, *Ibid.*, **192**, 16.
COULSON, C. A. and RUSHBROOKE, G. S., 1940, *Proc. Camb. Phil. Soc.*, **36**, 193.
JACOBS, Miss J., 1949, *Proc. Phys. Soc.* (in press).
MOFFITT, W. E. and COULSON, C. A., 1948, *Proc. Phys. Soc.*, **60**, 309.

CVIII. *Notices of New Books and Periodicals received.*

The Quantum Theory of Fields. By G. WENTZEL. [Pp. x+224.] (New York and London : Interscience Publishers, 1949.) Price 36s. net.

ENGLISH speaking physicists will welcome the translation of Wentzel's excellent "Einführung in die Theorie der Wellenfelder", first published by Franz Deuticke, Vienna, in 1943. Compared with the original edition only minor changes in the text have been introduced. It is an extremely useful text-book for everyone who wants to master the basic knowledge necessary for the understanding of the modern theories of mesons and the recent developments in quantum electrodynamics.

In accordance with the didactic purpose of the book, the formal rather than the problematic aspect of the theory is emphasized. The book does not give an up-to-date account of the attempts which have been made to solve the fundamental difficulties of the quantum theory of interacting fields, but the reader will find a chapter dealing with Dirac's λ -limiting process. The book contains useful references to the original literature.

D. P.

The Physics of Rubber Elasticity. By L. R. G. TRELOAR. [Pp. 252.] (Oxford: University Press.) Price 21s.

THIS contribution to the series of monographs on the physics and chemistry of materials will be of interest to many physicists and chemists not directly concerned with the mechanical behaviour of long-chain molecules. Dr. Treloar bases his exposition on the statistical theory of free energy changes in deformed networks of chains. The first part of the book is devoted to this subject, ending with two chapters on modifications of the theory at large strains, and the deviations of real rubbers from theoretically expected behaviour.

After a mainly experimental account of further points in the optical and dynamical behaviour of rubbers, the book ends with a chapter on the mathematical discussion of large elastic strains.

A particularly valuable aspect of the work is the clear way in which the points at which the theory remains unsatisfactory are indicated.

The Physical Society Reports on Progress in Physics. Vol. 2 (1948-49). Price £2 2s.

THE Physical Society has been remarkably successful in obtaining excellent articles on subjects of current interest from authors of eminence; the editor of this volume must have exceptional persuasive powers to have persuaded so many outstanding scientists to produce articles all at the same time. To select some from the list of articles would be invidious, and the present reviewer would prefer to confine himself to a list, which is given below:—

- "Mass Spectrometry." By H. G. Thode and R. B. Shields.
 - "Nuclear Paramagnetism." By B. V. Rollin.
 - "Phosphors and Phosphorescence." By G. F. J. Garlick.
 - "Recent Nuclear Experiments with High Voltage X-Rays." By W. Bosley and J. D. Craggs.
 - "Linear Accelerators." By D. W. Fry and W. Walkinshaw.
 - "Viscosity and Related Properties in Glass." By G. O. Jones.
 - "Theory of Oxidation of Metals." By N. Cabrera and N. F. Mott.
 - "Fracture and Strength of Solids." By E. Orowan.
 - "Multipole Radiation in Atomic Spectra." By A. Rubinowicz.
 - "Collisions between Atoms and Molecules at Ordinary Temperatures." By H. S. W. Massey.
 - "Low Temperature Physics." By K. Mendelssohn.
 - "Slow Neutron Absorption Cross Sections of the Elements." By M. Ross and J. S. Story.
 - "Molecular Distribution and Equation of State of Gases." By J. de Boer.
- N. M.

Luminescent Materials. By G. F. J. GARLICK. [Pp. viii+254.] (Oxford University Press, 1949.) Price 21s.

THIS book, one of the series of Monographs on the Physics and Chemistry of Materials, is devoted mainly to the study of the behaviour of inorganic phosphors excited by light. After an introduction in which he states the general conditions of luminescence, the author describes the theoretical model of a phosphor and derives the phosphorescence and thermoluminescence laws that may be expected to hold under various conditions. Next comes a review of the properties of the main types of phosphors; this is followed by a summary of the results of more recent work on luminescence efficiency, thermoluminescence, photoconductivity, dielectric constant and the effects of infra-red radiation, from which some more accurate information on the nature, distribution and action of the luminescence centres and electron traps can be obtained. Luminescence under electron bombardment (for the study of which some improvements in experimental technique are desirable) and the luminescence of organic molecules are dealt with in the last two chapters. In a synopsis the author puts forward some suggestions for further studies. The book is well documented and clearly written, and should appeal to everyone interested in its subject.

R. B.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

*CIX. A Technique for Rendering Approximate Solutions to
Physical Problems Uniformly Valid*.*

By M. J. LIGHTHILL †.

[Received August 22, 1949.]

SUMMARY.

A method is described for treating some of the characteristically non-linear problems of physics, in particular those involving a non-linear partial differential equation for which an approximate linearization is permissible everywhere except in a limited region, such as the neighbourhood of (§5) a singular characteristic of the approximate solution, or of (§6) the point at infinity, where the approximation is valueless. The method involves a transformation of an independent variable, which is determined progressively with successive approximations to the solution: only one step being necessary if a first approximation valid uniformly (even in the critical region) is to be obtained. The method is most easily understood in its application to simple first order ordinary differential equations, which are studied in detail in §§2 and 3 as a preparation for the extension to more complicated problems in §§4, 5 and 6. Physically, the longest section, §6, concerns the "spread" of a progressive wave at infinity, an important and essentially non-linear process.

§1. INTRODUCTION.

A TECHNIQUE will be described which the author has found useful for improving an approximate solution to a physical problem so as to make it valid uniformly in the range of the variables which is of physical interest. The problem will usually consist of an ordinary or partial differential equation to be solved under given boundary conditions when a certain parameter is small. The author has been concerned with special problems of this kind for some years and has used other methods in dealing with them previously (Lighthill 1948, and unpublished work): but the method to be described is the only one he has found which can be applied in any degree of generality. He wishes to thank Mr. G. B. Whitham for his collaboration in discovering the method, especially in its application to

* Expansion of a lecture given to the London Mathematical Society on May 19th, 1949.

† Communicated by the Authors.

the problem described in §6, and also Professor Goldstein for pointing out its similarity to the method of finding the limit cycle of a self-excited oscillatory system, which goes back to Poincaré (see §4). He expects that other authors may have used such an approach in special problems, but foresees advantage from the following general description.

The method is rather too widely applicable to be reduced to a theorem, on equations of a particular type. The author rather hopes to communicate ideas to fellow-scientists, which will help them to solve problems in their own particular fields. Another obstacle to a theorem is that he is unable to give a general proof of the validity of the answers afforded. Though for some special problems he has published *ad hoc* proofs, in others it would seem almost impossible to find one. Of course, in most physical theories there is some element of uncertainty already present, relating to the physical assumptions, and therefore physicists do not often spend time in proving results about whose truth they are once convinced. They may be convinced by physical considerations: or, for example, if they have a power series solution of a type that may be expected to have a radius of convergence, they may be prepared to accept its uniform convergence within *some* radius in a given range of some parameter, if each coefficient separately remains bounded in this range—a principle which will be appealed to in what follows. However pure mathematicians are invited to attempt a rigorous investigation, which would naturally be welcome.

The problems considered will be non-linear. The methods available for solving linear differential equations are very numerous. But a large number of problems in classical physics, which are of practical importance and interest to-day, involve non-linear differential equations. In much of the newer physics we make up our own equations and so far nobody has been bold enough to make these non-linear. However, it has been suggested from various quarters that this may finally be necessary, which provides a further incentive to find out all that is possible about such problems.

There are only a few methods available for treating non-linear equations: only methods of reducing them by approximation to one or more linear equations will be discussed here. This is done by assuming that some quantity, say α , is small— α may be a variable or a parameter—and expanding the unknown u , and the whole equation, in powers of α , perhaps after a simple transformation. In many cases the equations derived as the coefficients of powers of α are a set of linear equations in the unknown coefficients of the power series for u . Of the resulting series solution the investigator may use only the first partial sum, neglecting order α , or the second, neglecting order α^2 , etc. An alternative approach, by successive approximations, is essentially similar.

These methods are very familiar as part of the stock-in-trade of applied mathematics. It is not usually necessary, and frequently would be impossible, to investigate the convergence of the process: it is assumed

that it has some radius of convergence, and the computation of early terms may give weight to that assumption. However in certain important cases it is seen that *the radius of convergence becomes smaller and tends to zero as some other variable, say x , tends to a limit x_0 , finite or infinite*. In this case no partial sum is a good approximation near this critical value, and the approximations, though valid at each point (before the critical value is reached) when α is small enough, are not uniformly so as $x \rightarrow x_0$. The state of affairs is usually diagnosed by the fact that the ratio of the coefficients of (at any rate some pairs of) successive powers of α tends to infinity as $x \rightarrow x_0$: it is almost always heralded by some sort of singularity of the first approximate solution at $x=x_0$. It is usually particularly important to know an approximate solution near this point x_0 (this is sometimes because we have to apply a boundary condition near this point).

By the technique of this paper the series solution for the unknown u can be improved so as to make it valid uniformly in the neighbourhood of $x=x_0$. The idea is really very simple, though the fact that it achieves the object is more complicated. It consists of expanding not only the unknown u in powers of α , but also expanding one of its arguments (the independent variables), say y , in powers of α . In many cases y will simply be x . Of course y is technically independent, so to effect this expansion a new undefined variable z must be introduced and both u and its argument y expanded in powers of α with coefficients functions of z . Without loss of generality one may take $y=z$ on a first approximation: then the form of the expansions is

$$\left. \begin{aligned} y &= z + \alpha y_1(z, \dots) + \alpha^2 y_2(z, \dots) + \dots, \\ u(y, \dots) &= u_0(z, \dots) + \alpha u_1(z, \dots) + \dots, \end{aligned} \right\}, \quad \dots \quad (1)$$

where the dots following y, z in both equations indicate other variables on which u may depend. The ordinary expansion would consist of this second equation alone, with y for z . The additional set of arbitrary functions y_1, y_2, \dots are now determined so as to make both power series uniformly convergent in the required range. z is then determined implicitly in terms of y by the first of equations (1). This implicitness is very important as implying the possibility of branch-points of u as a function of y , which occur near $x=x_0$ in certain cases and can be determined approximately by the present technique as will be seen.

Consider the case of an ordinary differential equation for u in terms of x , with α a parameter. The usual expansion would be

$$u = u_0(x) + \alpha u_1(x) + \dots, \quad \dots \quad (2)$$

where $u_0(x), u_1(x), \dots$ satisfy linear equations, but if this has a radius of convergence tending to zero somewhere the following new form is suggested,

$$u = u_0(z) + \alpha u_1(z) + \dots, \quad x = z + \alpha x_1(z) + \alpha^2 x_2(z) + \dots, \quad \dots \quad (3)$$

where the functions displayed remain to be determined, after which the second equation will define the new variable z . The differential equation must be linear when $\alpha=0$: thus, if it is of the first order its form may be

$$[p(x) + \alpha p_1(x, u) + \dots] \frac{du}{dx} + q(x)u = r(x) + \alpha r_1(x, u) + \dots, \quad (4)$$

and most frequently the singularity described above is due to the fact that *the linearized coefficient of the highest derivative occurring vanishes at* $x=x_0$: for equation (4), that $p(x_0)=0$. This means that suddenly the next terms, here $\alpha p_1(x, u) du/dx$, become supremely important in this neighbourhood, even to a first approximation.

If the zero of $p(x)$ at $x=x_0$ is simple, divide (4) by the regular function $p(x)/(x-x_0)$ and move the origin to x_0 . (This transformation is not strictly necessary for the application of the technique and is done here principally to simplify the calculations in this paper.) An equation is derived with $x du/dx$ as its leading term. To simplify the calculations still further, only the equation

$$(x + \alpha u) \frac{du}{dx} + q(x)u = r(x), \quad (5)$$

which typifies the behaviour encountered, is here considered; but in physical problems there will usually be additional terms of order α or smaller, dependent on x and u , both in the coefficient of du/dx and on the right-hand side. These complicate but do not qualitatively alter the method, unless they are terms which happen to become unduly large.

In fact equation (5), which is considered in great detail in §§2 and 3, gives an excellent introduction to the whole subject; and the behaviour when there are additional terms, or when (§4) the equation is of higher order but the singularity is still a regular one of its linearized form, or when (§5) it is partial and the non-uniformity appears as a certain characteristic curve of the linearized equation is approached, follows closely analogous lines.

An equation whose linearized form has an irregular singularity will also be discussed in §4; and in §6 a singularity of another type, possible only with a partial differential equation, due to the gradual divergence at infinity of its characteristics from their approximate forms, will be treated by the method of this paper.

Particular importance attaches to §§5 and 6, since far less is known about non-linear partial differential equations than about ordinary ones (for which methods based on inequalities, geometry or topology may prove much). The work of §5 will enable problems involving characteristics of discontinuity, or (worse) shock waves, to be solved, and these curves determined, to any degree of approximation. That of §6 enables non-linear wave motion to be studied, with equally good approximation at all distances from the initial region.

§2. FIRST ORDER ORDINARY EQUATION WHEN THE APPROXIMATE SOLUTION TENDS TO INFINITY.

Substitution of the expansions (3) in (5) gives

$$\begin{aligned} & (z + \alpha x_1 + \alpha^2 x_2 + \dots + \alpha u_0 + \alpha^2 u_1 + \dots)(u'_0 + \alpha u'_1 + \alpha^2 u'_2 + \dots) \\ &= (1 + \alpha x'_1 + \alpha^2 x'_2 + \dots)[r + \alpha x_1 r' + \alpha^2 x_2 r' + \frac{1}{2} \alpha^2 x_1^2 r'' + \dots \\ & \quad - (q + \alpha x_1 q' + \alpha^2 x_2 q' + \frac{1}{2} \alpha^2 x_1^2 q'' + \dots)(u_0 + \alpha u_1 + \alpha^2 u_2 + \dots)], \quad (6) \end{aligned}$$

where r, q stand for $r(z), q(z)$ and primes denote differentiation with respect to z . The term independent of α in (6) is

$$zu'_0 + qu_0 = r, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

namely the straightforward linearization of (5), with its solution derived from the equation

$$\frac{d}{dz} \left(u_0 \exp \int \frac{q}{z} dz \right) = \frac{r}{z} \exp \int \frac{q}{z} dz. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

The functions $q(z), r(z)$ are assumed regular at $z=0$. If $q(0)=q_0$, then $\exp \int (q/z) dz = z^{q_0} \mathbf{R}$, where *throughout this paper* \mathbf{R} stands for any function of z regular at $z=0$. It is easily deduced from (8) that $u_0 = \mathbf{R} + O(z^{-q_0})$ as $z \rightarrow 0$, except that if q_0 is a non-positive integer then $u_0 = \mathbf{R} + O(z^{-q_0} \log z)$.

The coefficient of α in (6) gives the equation

$$\frac{d}{dz} \left(u_1 \exp \int \frac{q}{z} dz \right) = \frac{1}{z} \exp \int \frac{q}{z} dz [(r - qu_0)x'_1 + (r' - q'u_0 - u'_0)x_1 - u_0 u'_0]. \quad (9)$$

If the ordinary form of expansion had been used, with $x=z$, that is $x_1=0$, we would deduce from (9) that

$$u_1 = O(z^{-q_0}) + O(u_0 u'_0) = \mathbf{R} + O(z^{-q_0-1}) + O(z^{-2q_0-1});$$

(except that when q_0 is a non-positive integer $u_1 = \mathbf{R} + O(z^{-q_0-1} \log z)$.) The precise effect of these singularities in u_1 depends on the sign of q_0 , but in any case they are disastrous. Thus if $q_0 > 0$ the ratio of successive terms u_0 , of order z^{-q_0} , and αu_1 , of order αz^{-2q_0-1} , in the power series for u is $O(\alpha/z^{q_0+1})$; and it is found by continuing the process that this ratio persists, giving a radius of convergence which tends to zero like $O(x^{q_0+1})$ as $x \rightarrow 0$. But if $q_0 < 0$ the term $O(z^{-q_0-1})$ in u_1 is more serious than the term $O(z^{-2q_0-1})$, since in u_2 it produces an $O(z^{-q_0-2})$ term, and ultimately negative powers of z appear (with logarithmic multipliers if q_0 is an integer), leading again to divergence for given α when $x=z$ is small enough.

To give u_1 only a similar singularity to that of u_0 , x_1 would have to be chosen so that the factor in square brackets in (9) is $\mathbf{R} + O(z^{-q_0+1})$. This can be done: in fact x_1 can be chosen to make the factor vanish if required; but only a finite number of terms in an ascending power series for x_1 are really needed. The case $q_0 < 0$, when the first approximate solution u_0 remains bounded, will be postponed to §3. At present suppose that $q_0 > 0$.

If $u_0 \sim Az^{-q_0}$ as $z \rightarrow 0$ (the constant A will be different for different solutions of (7) and will be determined by some boundary condition) then the solution for x_1 required to render the factor in square brackets zero, or at least innocuous, must satisfy

$$(-q_0Az^{-q_0} + \dots)x'_1 + (q_0Az^{-q_0-1} + \dots)x_1 = -A^2q_0z^{-2q_0-1} + \dots \quad (10)$$

or $x_1 \sim -Az^{-q_0}/(q_0+1)$ as $z \rightarrow 0$.

In the same way the behaviour of $x_j(z)$ as $z \rightarrow 0$ which is necessary to make $u_j = O(z^{-q_0})$ as well as u_0, u_1, \dots, u_{j-1} , can be investigated by induction. Assume that $x_k(z) = O(z^{-kq_0})$ (which is already proved for $k=1$) when $k < j$. The condition on u_j requires that a factor of the form

$$(r-qu_0)x'_j + (r'-q'u_0-u'_0)x_j + \text{terms in } q, r, u_0, u_1, \dots, u_{j-1}, x_1, x_2, \dots, x_{j-1} \quad (11)$$

should vanish, or at least be $R + O(z^{-q_0+1})$. The terms unspecified in (11) are the terms independent of x_j or u_j in the coefficient of α^j in the right-hand side of (6) minus the left-hand side. Of these the ones of largest order as $z \rightarrow 0$ are seen to be those of the form $-x'_k u_0$ times the coefficient of α^{j-k} in $q(x)$ when x is expanded as $z + \alpha x_1 + \alpha^2 x_2 + \dots$. These are $O(z^{-kq_0-1} \cdot z^{-q_0} \cdot z^{-(j-k)q_0}) = O(z^{-jq_0-q_0-1})$. Hence x_j (which is multiplied by $-u'_0$ and its derivative by $-qu_0$ in (11), these two products being of the same sign) is $O(z^{-jq_0})$ as $z \rightarrow 0$, which completes the induction. Thus it has been shown that

$$x = z + \alpha \left(-\frac{Az^{-q_0}}{q_0+1} + \dots \right) + \alpha^2 O(z^{-2q_0}) + \alpha^3 O(z^{-3q_0}) + \dots \quad (12)$$

In accordance with the principle enunciated in §1 it is induced that the radius of convergence ρ of the power series (12) for x is of order z^{q_0} as $z \rightarrow 0$. This slight improvement, over the radius $O(z^{q_0+1})$ for the series for u when z is taken equal to x , seems at first sight almost worthless: *but in fact it is crucial*. To see this first observe that, by (12), $x=0$ itself corresponds to a value of z of order $\alpha^{1/(1+q_0)}$ so that at this point ρ is of order $\alpha^{q_0/(1+q_0)}$ and hence exceeds α when α is small enough, giving convergence even at the singularity itself for small enough α . But not only is $x=0$ an interior point of the region of convergence (for a suitably small value of α) but the boundaries of this region are at a distance from $x=0$ independent of α . In fact if z is of order α^{1/q_0} but $|\alpha|$ is less than ρ , then by (12) x will have a value with a finite limit (in general not zero) as $\alpha \rightarrow 0^*$. This region of validity of the double expansion (3), namely one independent of α (though unspecified) and including $x=0$, is the best that could possibly be obtained on the assumptions made—which do not exclude singularities of the equation, for example singularities of $q(x)$ or $r(x)$, at a finite distance from the origin.

* And these limits fill a region obtained by conformal transformation from a circle in the αz^{-q_0} plane, whose centre (the origin) corresponds to $x=0$.

The roughest approximation which is valid uniformly near $x=0$, that is the roughest which has *any* value near the origin, is

$$u=u_0(z), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

where z is the root of $x=z-\alpha A z^{-q_0}/(q_0+1)^*$ which is approximately x when $x \gg 0$.

(Here it is assumed that the solution "starts" from a positive value of x , from which it is required to continue it downwards and through $x=0$.) There is a branch point (of order two) of z as a function of x , and so also of u as a function of x (in general), at a point given by $dx/dz=0$, that is by

$$z \doteq \left(-\frac{\alpha A q_0}{q_0+1}\right)^{1/(q_0+1)}, \quad x \doteq \left(1+\frac{1}{q_0}\right) \left(-\frac{\alpha A q_0}{q_0+1}\right)^{1/(q_0+1)} \quad (14)$$

It is important to notice that this branch-point (to whose position (14) gives a first approximation) represents a real singularity of the exact solution of (5), not an artificial one due to the approximations adopted: its cause is in fact the vanishing of the coefficient $x+\alpha u$ of du/dx , which may be verified to a first approximation as follows:†

$$\left(1+\frac{1}{q_0}\right) \left(-\frac{\alpha A q_0}{q_0+1}\right)^{1/(q_0+1)} + \alpha A \left[\left(-\frac{\alpha A q_0}{q_0+1}\right)^{1/(q_0+1)} \right]^{-q_0} = 0. \quad (15)$$

The constant $A=\lim z^{q_0}u(z)$ is different for different solutions of the equation. If $A<0$ the branch-point is for a real positive value of x . In a physical problem the solution will normally "stop" at or before this point; though as a solution of the complex variable problem the two series (3) are still valid in a neighbourhood of $x=0$ (on a Riemann surface). But if $A>0$ *no branch-point appears for real x* , and the solution is regular for real x within a distance of the origin independent of α . This is because $x+\alpha u$ remains positive even when x becomes negative. It is interesting that an approximate solution is still possible although u becomes large enough for this to be so.

In concluding the general discussion of the case $q_0>0$ it must be remarked that, as is now seen, it would have been permissible to choose the functions $x_j(z)$ so that, for each j , $u_j(z)=O(z^{-(j+1)q_0})$, not simply $=O(z^{-q_0})$. (For it has been seen that a radius of convergence of order z^{q_0} is permissible; and the reader will verify that the proof by induction that $x_j(z)=O(z^{-jq_0})$ is unaffected in its conclusion by the change, though terms of order $O(z^{-jq_0-q_0-1})$ will appear, among those unspecified in (11), additional to

* For all z , the terms here picked out from (12) exceed (in order, as $\alpha \rightarrow 0$) the remainder.

† In the (x, u) plane the branch-point is somewhere on the line joining the origin to the saddle-point singularity of the equation, given by $x+\alpha u=0$, $q(x)u=r(x)$. The presence of a branch-point near such a singularity in a first order equation follows from the general theory of such singularities: however, the extension of the present method to higher order equations, though immediate, produces analogous results which could not be treated by the ordinary theory (see § 4).

those already described in the proof: these have the form $-x_k u'_{j-k}$, $-u_{k-1} u'_{j-k}$, or $-x_k u_l$ times the coefficient of α^{j-k-l} in $q(x)$. This fact shows that only a very few terms in an ascending series for x_j are required (for example, one only in x_1 if $q_0 \geq 1$).

Example. The reader will understand the method most clearly if he applies it now to solve the equation

$$(x + \alpha u) du/dx + (2+x)u = 0 \quad . \quad . \quad . \quad (16)$$

under the boundary condition $u = e^{-1}$ when $x = 1$. (He will easily verify that this implies the conditions $u_0(1) = e^{-1}$, $u_1(1) = u'_0(1)x_1(1)$, . . .) After some calculation he will find as the first terms of the series (3) for u and x

$$\left. \begin{aligned} u &= e^{-z} z^{-2} + \alpha \left[e^{-z} z^{-2} \left\{ \frac{2}{3z^3} + \frac{1}{3z^2} - \int_z^1 e^{-z} \left(\frac{2}{z^4} + \frac{1}{z^3} \right) dz \right\} \right] + O\left(\frac{\alpha^2}{z^6}\right), \\ x &= z - \frac{\alpha}{3z^2} - \frac{3\alpha^2}{10z^4} + O\left(\frac{\alpha^3}{z^6}\right). \end{aligned} \right\} \quad (17)$$

At $x = 0$, $u = (3/\alpha)^{\frac{1}{3}} - 3 \cdot 9(3/\alpha)^{\frac{1}{3}} + O(1)$. Probably in this example the series (17) are valid in any bounded interval of x for small enough α , since the origin is the only singularity of the linearized equation, and no branch-point appears for real x .

An exception to the foregoing results will occur when $A = \lim_{z \rightarrow 0} z^{q_0} u_0(z)$ is zero. In this case $x_1(z)$ may be taken identically zero and $u_1 = \mathbf{R} + O(z^{-q_0})$ will still hold. In general—that is, if $\lim_{z \rightarrow 0} z^{q_0} u_1 \neq 0 - a$ value of $x_2(z)$ which is of order z^{-q_0} as $z \rightarrow 0$ will now be necessary to prevent excessive growth of u_2 ; and successive x_j after this will as before bear a ratio of order z^{-q_0} to one another, giving a set of qualitative phenomena in the complete solution similar to those observed before. (For example the branch-point will now be at a distance of order $\alpha^{2/(q_0+1)}$ from the origin.) But there will be one special solution of the differential equation (5) for which every one of the limits $\lim_{z \rightarrow 0} z^{q_0} u_j(z)$ is zero, and for this solution no functions $x_j(z)$ are necessary and it is permissible to take $x = z$. This single solution is bounded near the origin *uniformly* in α , and no singularity of it occurs near the origin even in the complex plane.

To conclude this section a brief account of the case $q_0 = 0$ is given. In this case $u_0 = r_0 \log z + A + O(z \log z)$ as $z \rightarrow 0$, where A is arbitrary, and it is found that if $x_1 = -r_0 \log z - (r_0 + A)$, and generally if x_j is a certain expression which is $O(\log^{2j-1} z)$, then these expressions can be chosen so that u_j is $O(\log^{2j+1} z)$. The radius of convergence is of order $(\log z)^{-2}$, where the point $x = 0$ is given by the approximate equation

$$0 = z - \alpha[r_0 \log z + r_0 + A],$$

whence

$$z = O(\alpha \log \alpha).$$

Hence ρ is of order very considerably greater than α at $x=0$. The branch-point is at $z=\alpha r_0+O(\alpha^2 \log^2 \alpha)$, $x=-\alpha r_0 \log(\alpha r_0)-\alpha A+O(\alpha^2 \log^3 \alpha)$, which is real and positive (for small α) if $r_0>0$. If $r_0<0$ no branch-point occurs for real x : any finite real value of x is given by putting some positive value of z into the expression $z-\alpha[r_0 \log z+r_0+A]$, and the series solution is presumably valid up to the neighbourhood of any singularities of the approximate equation other than zero.

§3. FIRST ORDER ORDINARY EQUATION WHEN THE APPROXIMATE SOLUTION REMAINS BOUNDED.

The argument which will be applied for negative q_0 is not as complete as that of §2. It will first be given and then criticized. It is convenient to consider the case $q_0\leq -1$ first; when in u_1 (for example), of the encroaching terms of orders z^{-q_0-1} and z^{-2q_0-1} (*vide post* equation (9)), only the first is important, the other being already $O(z^{-q_0})$. In this case it is found to be possible to take each $x_j(z)$ as a *constant*, in such a way that the series for u is uniformly convergent in some α -circle for all $z\geq 0$. The point $z=0$ is a singularity of every solution, corresponding to a value of x at a distance $O(\alpha)$ from the origin; for x only differs from z by a constant (depending on α).

These facts can be verified directly, as were the corresponding results in §2. More simply they can be verified by making the preliminary transformation $x=z+\epsilon$, $u=v+\delta$, where ϵ , δ are constants depending on α so chosen that in the equation for v in terms of z both the coefficient of dv/dz and the right-hand side vanish when $v=z=0$. Such constants can be determined for any equation of the general form (4). For the specially simple equation (5) they must satisfy $\epsilon+\alpha\delta=0$ and $r(\epsilon)=\delta q(\epsilon)$. The solutions of these equations can be expanded in a regular power series in α (by Weierstrass's inversion theorem) if r , q are regular at the origin: for example

$$\epsilon=\alpha(-r_0/q_0)+\alpha^2(r_0(r_0q_1-r_1q_0)/q_0^3)+\dots$$

In the new equation dv/dz is equated to the quotient of two functions of v and z each regular and vanishing at the origin, which is a nodal singularity of the equation into which all solutions pass. The author contents himself with showing that for the simplified equation of this type,

$$(z+\alpha v)\frac{dv}{dz}+q(z)v=r(z), \quad r_0=0, \quad \dots \dots \dots (18)$$

when $q_0\leq -1$, a uniformly convergent power series in α for v with coefficients functions of z is possible. For substitution of the expansion $v=v_0(z)+\alpha v_1(z)+\dots$ in the equation gives

$$v_0(z)=\exp\left(-\int\frac{q}{z}dz\right)\int\left(\frac{r}{z}\exp\int\frac{q}{z}dz\right)dz=z\mathbf{R}+O(z^{-q_0}\log^\mu z), \quad (19)$$

where μ is 1 when q_0 is an integer and otherwise zero, and

$$\frac{d}{dz} \left(v_j \exp \int \frac{q}{z} dz \right) = -\frac{1}{z} \exp \int \frac{q}{z} dz \left[v_0 v'_{j-1} + v_1 v'_{j-2} + \dots + v_{j-1} v'_0 \right]. \quad (20)$$

Assume that for $k < j$, as has been shown for $k=0$, $v_k = z\mathbf{R} + O(z^{-q_0} \log^{v_k+\mu} z)$, where $v=2$ when $q_0=-1$ and $v=1$ otherwise. Then the square bracket in (20) is of the form $z\mathbf{R} + O(z^{-q_0} \log^{v(j-1)+\mu} z)$, which on multiplication by $z^{-1} \exp \int (q/z) dz$, and integration, gives $z^{q_0+1}\mathbf{R} + O(\log^{vj+\mu} z) + \text{constant}$, showing that $v_j = z\mathbf{R} + O(z^{-q_0} \log^{vj+\mu} z)$ and completing the induction. In particular each v_j is bounded as $z \rightarrow 0$ and the convergence therefore is assumed to be uniform in $z \geq 0$. Each solution has a complicated singularity at $z=0$, that is at $x=\epsilon$: the roughest approximation of value for small x is $u=v_0(x-\epsilon)+\delta$.

But here the author fears that the application of the principle of §1 may have led him into error. This is because, as has just been shown, in the series $\sum \alpha^k v_k(z)$, if the regular parts of the v_k be discarded, the remaining series has coefficients each greater than the last by a factor of order $\log z$. Thus it seems possible that the radius of convergence tends to zero like $|\log z|^{-\nu}$ as $z \rightarrow 0$, where ν is 2 when $q_0=-1$ and 1 otherwise. It is therefore possible that for given α a singularity occurs before the point $z=0$ (where x is given by the rigorously convergent series $\sum_{k=1}^{\infty} \alpha^k x_k = \epsilon$); but its distance from $z=0$ must be $O(\exp(-\text{const.}/\alpha^{1/\nu}))$, which is smaller than any power of α .

Such a singularity can occur. Thus the exact solution of the equation $(z+\alpha v) dv/dz - v=0$ with the boundary condition $v(1)=1$ is given by $z=v(1+\alpha \log v)$. The function $v(z)$ has a branch-point when $dz/dv=0$, or $z=-\alpha \exp(-1-\alpha^{-1})$. But the author knows of no example with $q_0 \neq -1$ and believes that the phenomenon perhaps occurs only for this value: thus it is only for this value that the order as $z \rightarrow 0$ of $v_k(z)$ itself grows with k , or that the differentiated series has unbounded coefficients. But in any case the above work has narrowed down the region, in which an approximate solution (other than $v=0$) is *not* known, to an exponentially small *open* interval (for at $z=0$ it must vanish): the author knows no general method of elucidating matters in this interval, though he did so rigorously in a special case with $q_0=-1$ (Lighthill 1948).

In the range $-1 < q_0 < 0$ the characteristic properties of the ranges $q_0 \leq -1$ and $q_0 \geq 0$ are combined. Thus the problem is simplified by the transformation used above: the equation in the form (17) will be considered here. (Again in a practical problem no such transformation is necessary: the expansions can be substituted directly into the given equations. The transformation is adopted to give the reader as little trouble as possible in following how and why the method works.) But even the transformed equation cannot be solved by a single series for $v(z)$ when $-1 < q_0 < 0$. For then a term of order z^{-2q_0-1} appears in v_1 , and similarly a term of order $z^{-(j+1)q_0-j}$ in v_j , which is unbounded for large j , giving non-uniform convergence as $z \rightarrow 0$.

A double expansion, as before, is needed to overcome the difficulty. To uniformize the notation suppose the (in general already transformed) equation written in the form (5) with variables u and x , but with $r_0=0$ and $-1 < q_0 < 0$. Substituting the double expansion (3) in the equation, it is necessary to choose $x_1(z)$, $x_2(z)$, . . . , to eliminate the terms of order larger than $z^{-q_0} \log^j z$ in u_j : these are terms of order z^{-2q_0-1} (but not z^{-q_0-1} , since $r_0=0$) in $u_1(z)$, etc. It is found sufficient as in §3 to take $x_1(z) = -Az^{-q_0}/(q_0+1)$, and for the higher terms $x_k(z) = O(z^{-q_0} \log^{k-1} z)$, $u_k(z) = O(z^{-q_0} \log^k z)$. For assuming these results for $k < j$, the largest terms unspecified in (11) occur in products of the form $x_k u'_{j-k}$ or $u_{k-1} u'_{j-k}$ or $x'_k x_{j-k} r'$ or $-x'_k q u_{j-k}$, which are all $O(z^{-2q_0-1} \log^{j-1} z)$. The whole expression (11) can therefore be rendered $O(z^{-q_0} \log^{j-1} z)$ by an x_j which is $O(z^{-q_0} \log^{j-1} z)$; since, in (11), x_j is multiplied by a term in z^{-q_0-1} from $-u'_0$ and x'_j by a term in z^{-q_0} from $-qu_0$, these two products being of opposite sign but the latter being asymptotically smaller as $z \rightarrow 0$. It has been shown that

$$\frac{d}{dz} \left(u_j \exp \int \frac{q}{z} dz \right) = \frac{1}{z} \exp \int \frac{q}{z} dz [O(z^{-q_0} \log^{j-1} z)], \quad . \quad (21)$$

and it now follows as in the case $q_0 \leq -1$ that $u_j = O(z^{-q_0} \log^j z)$, completing the induction. The power series for both u and x have bounded coefficients in $z \geq 0$. Thus on the principle of §1 they are uniformly convergent: but on the basis of the criticisms of the last paragraph but two they may have a radius of convergence of order $|\log z|^{-1}$ as $z \rightarrow 0$, and so for given α be valid only outside a neighbourhood of $z=0$ of magnitude $O(\exp(-\text{const.}/\alpha))$.

The roughest approximation which has any value near the origin is again given by equation (13); and a branch-point of z (and so also of u) as a function of x will still be given by equation (14). Here A still signifies $\lim_{z \rightarrow 0} z^{q_0} u_0(z)$. Though $x=0$ (where $u=0$) is still (as when $q_0 \leq -1$) a nodal singularity, it may (since there are branch-points) be on a different sheet of the Riemann surface from that "on which one starts"; the origin on the latter sheet being no singularity at all. For example, confining oneself to *real* x , there is now a branch-point (14) only if $A > 0$, and at it x is negative. When x exceeds this value the solution is regular, and α is certainly within its radius of convergence. The origin on this sheet corresponds to a regular point $z = (\alpha A / (q_0 + 1))^{1/(q_0+1)}$; and the nodal singularity $u=x=0$ is reached only after the branch-point has been encircled once. The physical solution will "stop" at the branch-point and never reach the nodal singularity. But if $A < 0$ the latter is the only singularity for real x (unless perhaps there is another exponentially close).

Example.— $(x + \alpha u) du/dx - \frac{1}{2} u = 1 + x^2$; $u(1) = -1$. The preliminary transformation is $x = z + \epsilon$, $u = v - (\epsilon/\alpha)$, where

$$\epsilon = (1 - \sqrt{(1 - 16\alpha^2)})/4\alpha = 2\alpha + 8\alpha^3 + \dots$$

The complete series solution will be found to be

$$\left. \begin{aligned} u &= (-\epsilon/\alpha) + \frac{1}{3}z^{\frac{1}{2}} + \frac{2}{3}z^2 + \alpha \left(-\frac{21}{5}z^{\frac{1}{2}} + 8z - \frac{13}{9}z^{\frac{3}{2}} - \frac{16}{45}z^3 \right) + O(\alpha^2 z^{\frac{1}{2}} \log z), \\ x &= \epsilon + z - \frac{2}{3}\alpha z^{\frac{1}{2}} + \frac{42}{5}\alpha^2 z^{\frac{1}{2}} + O(\alpha^3 z^{\frac{1}{2}} \log z). \end{aligned} \right\} \quad \dots (22)$$

There is a branch-point at

$$z = \frac{1}{9}\alpha^2 - \frac{14}{5}\alpha^3 + O(\alpha^4 \log \alpha),$$

where

$$x = 2\alpha - \frac{1}{9}\alpha^2 + \frac{54}{5}\alpha^3 + O(\alpha^4 \log \alpha)$$

and $u = -x/\alpha$.

The nodal singularity $x = \epsilon$, $u = -\epsilon/\alpha$ is on a different sheet from the initial position $x = 1$, $u = -1$.

§4. SECOND ORDER ORDINARY DIFFERENTIAL EQUATIONS.

Problems involving differential equations of order higher than the first can normally, as far as the author has experimented, be treated by methods essentially similar to those of §§2 or 3 (according as the first approximate solution is unbounded or bounded), provided that the first approximate equation has only a regular singularity at the critical point (so that the singularity of the solutions in the neighbourhood is algebraic or logarithmic). It is strongly recommended for ease of treatment that the problems be expressed in the form of a system of first order equations.

For example the second order equation

$$\left(x + \alpha \frac{dv}{dx} + \alpha av \right) \frac{d^2v}{dx^2} + q(x) \frac{dv}{dx} + s(x)v = r(x), \quad \dots (23)$$

should be rewritten

$$(x + \alpha u + \alpha av) \frac{du}{dx} + q(x)u + s(x)v = r(x), \quad \frac{dv}{dx} = u, \quad \dots (24)$$

and it can then be treated almost exactly as in §§2 and 3, with the additional expansion $v = v_0(z) + \alpha v_1(z) + \dots$, satisfying

$$v'_0 + \alpha v'_1 + \dots = (1 + \alpha x'_1 + \alpha^2 x'_2 + \dots)(u_0 + \alpha u_1 + \dots). \quad \dots (25)$$

Little qualitative difference emerges from the extra terms $\alpha av \, du/dx + s(x)v$. Thus when $q_0 > 0$, and $u_0 \sim Az^{-q_0}$ as $z \rightarrow 0$, the roughest approximation for u valid near $x = 0$ is still (13): that for v is $v_0(z)$, determined from the equation $v'_0(z) = u_0(z)$ and some boundary condition. When $q_0 \leq -1$ the $x_k(z)$ may still (as in §3) be taken constant; but the constants can no longer be determined from the equation alone without the boundary

conditions*. The following rule is found: after $u_0(z)$, $u_1(z)$, ..., $u_{k-1}(z)$, $v_0(z)$, ..., $v_{k-1}(z)$ and x_1 , x_2 , ..., x_{k-1} have been determined, choose the constant x_k to make the coefficient of α^k in $x + \alpha u + \alpha v$ (or more generally in the coefficient of du/dx) zero at $z=0$. This can always be done as x_k appears linearly in this coefficient. Its effect is to prevent the successive increase in order by a factor z^{-1} in the non-regular parts of the u_k . When $-1 < q_0 < 0$ a combination of both approaches is in general necessary. Equations containing far more terms than (24) would often be amenable to the methods of this paragraph.

Example.—The waves set up by a cylinder expanding uniformly with radial velocity αx into still air, in which the velocity of sound is a , are bounded by a shock wave which expands with velocity αM , where M is a number whose excess over unity can be determined for small α by the present method. If the air's velocity at time t at a distance $\alpha t x$ from the axis of the cylinder is αu , then the equations

$$\left. \begin{aligned} [1 - x^2 + (\gamma + 1)xu - (\gamma - 1)v - \frac{1}{2}(\gamma + 1)u^2] \frac{du}{dx} \\ + \frac{u}{x} [1 + (\gamma - 1)(xu - v - \frac{1}{2}u^2)] = 0, \\ dv/dx = u, \end{aligned} \right\} \quad \dots \quad (26)$$

where γ is the adiabatic index, hold in $\alpha < x < M$ under the boundary conditions $u(\alpha) = \alpha$, $v(M) = 0$, $u(M) = 2(M - M^{-1})/(\gamma + 1)$. The first approximate solution is

$$u = \alpha^2 \sqrt{(x^2 - 1)}, \quad M = 1, \quad v = \int_1^x u dx, \quad \dots \quad (27)$$

with a singularity at $x=1$ corresponding to $q_0 = -\frac{1}{2}$. (This solution does not even exist when $x > 1$, while in fact u must exist in $x < M$.) This suggests the expansions

$$\left. \begin{aligned} u = \alpha^2 u_0(z) + \alpha^4 u_1(z) + \dots, \quad v = \alpha^2 v_0(z) + \alpha^4 v_1(z) + \dots, \\ x = z + \alpha^2 x_1 + \alpha^4 x_2 + \dots, \quad M = 1 + \alpha^2 M_1 + \alpha^4 M_2 + \dots, \end{aligned} \right\} \quad \dots \quad (28)$$

where $u_0(z) = \sqrt{(z^2 - 1)}$ and $v_0(z) = \int_1^z u_0(z) dz$. In x_1 , x_2 , ..., by the general theory with $-1 < q_0 < 0$, the most important terms will be constants, though terms in $(1-z)^{\frac{1}{2}}$, multiplied by various powers of $\log(1-z)$, will also be required. The coefficient of α^2 in the coefficient of du/dx in (26) is $-2zx_1 + (\gamma + 1)zu_0 - (\gamma - 1)v_0$, which vanishes at $z=1$ if $x_1=0$. With this value of x_1 the equation for u_1 is

$$(1 - z^2)u_1' + z^{-1}u_1 + [(\gamma + 1)zu_0 - (\gamma - 1)v_0]u_0' + z^{-1}u_0[(\gamma - 1)zu_0 - (\gamma - 1)v_0] = 0, \quad \dots \quad (29)$$

whence $u_1(1) = -(\gamma + 1) \lim_{z \rightarrow 1} (u_0 u_0') = \gamma + 1$. The value of $v_1(1)$ is obtained

* In fact in the (x, u, v) space there is no longer only one singular point as in §§ 2 and 3, but a line of them.

from the boundary condition :

$$\left. \begin{aligned} \text{When } z + \alpha^4 x_2 + \dots = M, \quad \alpha^2 v_0 + \alpha^4 v_1 + \dots = 0 \\ \text{and } \alpha^2 u_0 + \alpha^4 u_1 + \dots = \frac{2(M - M^{-1})}{\gamma + 1} \end{aligned} \right\} \dots \quad (30)$$

The latter condition, since $u_0 \sim \sqrt{2(1-z)}$ as $z \rightarrow 1$, shows that $M - 1 = O(\alpha^4)$, and the former that hence

$$v_1(1) = - \lim_{\alpha \rightarrow 0} [\alpha^{-2}(v_0)_{x=M}] = \lim_{\alpha \rightarrow 0} [\alpha^{-2}O(\alpha^6)] = 0.$$

Hence the value at $z=1$ of the coefficient of α^4 in the coefficient of du/dx in (26) is $-2x_2(1) + (\gamma + 1)^2$. By the general theory of §3, the value of $x_2(1)$ must be chosen to make this quantity zero. Hence finally, by (30), with $x_2 = \frac{1}{2}(\gamma + 1)^2 + O[(1-z)^{\frac{1}{2}} \log(1-z)]$,

$$\alpha^2 \sqrt{2(1-M) + \alpha^4(\gamma + 1)^2} + (\gamma + 1)\alpha^4 + O(\alpha^6 \log \alpha) = \frac{4(M-1)}{\gamma + 1} + O(M-1)^2, \quad (31)$$

whence $M = 1 + \frac{3}{8}(\gamma + 1)^2 \alpha^4 + O(\alpha^6 \log \alpha)$ measures the ratio of the speed of propagation of the disturbance to that of sound.

Presumably analogous methods would be applicable near a regular singularity at infinity of the first approximate equation: but near an irregular singularity matters are very different. For example the equation

$$\frac{d^2 u}{dx^2} + u = \alpha f\left(u, \frac{du}{dx}\right) \quad (32)$$

has ∞ as an irregular singularity of its approximate form. Self-excited oscillations, both mechanical and electrical, often satisfy equations of this form. The initial behaviour of the solution is an oscillation of period 2π and arbitrary amplitude, but the small disturbing term acting over a long time can alter this completely. Often its influence is to add energy whenever there is only a little and to subtract it when there is more than a certain amount, so that ultimately a stable oscillation sets in, known as a limit cycle. A method of finding the limit cycle which goes back to Poincaré is a special case of the technique of this paper.

Substitution of the expansions (3) in (32) gives

$$\begin{aligned} \frac{u_0'' + \alpha u_1'' + \dots}{(1 + \alpha x_1' + \dots)^2} - \frac{(u_0' + \alpha u_1' + \dots)(\alpha x_1'' + \alpha^2 x_2'' + \dots)}{(1 + \alpha x_1' + \dots)^3} + u_0 + \alpha u_1 + \dots \\ = \alpha f\left(u_0 + \alpha u_1 + \dots, \frac{u_0' + \alpha u_1' + \dots}{1 + \alpha x_1' + \dots}\right) \dots \quad (33) \end{aligned}$$

The term independent of α is $u_0'' + u_0 = 0$. A solution is $u_0 = A \sin z$, and the coefficient of α in (33) then becomes

$$u_1'' + u_1 = f(A \sin z, A \cos z) - 2Ax_1' \sin z + Ax_1'' \cos z. \quad (34)$$

When $z=x$, so that $x_1=0$, the right-hand side has period 2π . Hence $u_1 \rightarrow \infty$ as $z \rightarrow \infty$ (leading to divergence of the series for u), unless the two Fourier coefficients a_1 and b_1 in

$$f(A \sin z, A \cos z) = \frac{1}{2}a_0 + \sum_1^{\infty} (a_n \cos nz + b_n \sin nz)$$

vanish, in which case u_1 is periodic. This gives two conditions on A , which in general cannot be satisfied simultaneously. But if x_1 is taken as $b_1 z/2A$ it appears from (34) that the necessity for b_1 to vanish is eliminated. There remains one equation to determine the amplitude of a limit cycle, namely $a_1=0$ or

$$\int_0^{2\pi} f(A \sin z, A \cos z) \cos z \, dz = 0. \quad . \quad . \quad . \quad (35)$$

The period of the limit cycle is the change in x when z changes by 2π (since u_1 as well as u_0 is now periodic with period 2π), namely

$$2\pi(1 + \alpha b_1/2A + O(\alpha^2)),$$

or

$$2\pi + \frac{\alpha}{A} \int_0^{2\pi} f(A \sin z, A \cos z) \sin z \, dz + O(\alpha^2). \quad . \quad . \quad . \quad (36)$$

The process can be extended to further approximations with $x_j(z)$ always a multiple of z . However the question whether a technique of this form can be used to determine motions other than the limit cycle, or, in general, whether it can be recommended for use near an irregular singularity of the first approximate equation, remains unelucidated.

§5. PARTIAL DIFFERENTIAL EQUATION NEAR A SINGULAR CHARACTERISTIC OF THE APPROXIMATE SOLUTION.

In this section is studied the direct extension of the procedures of §2, §3 and the first part of §4 to problems of second order partial differential equations in two variables. In those sections the critical behaviour of the expansion of the unknown quantity (v in §4) was always associated with a singularity (algebraic or logarithmic) of the first approximate solution v_0 . Now when such a singularity occurs in the first approximate solution to a partial differential equation, it will be propagated along a characteristic of the (linear) equation for v_0 . This can be described as a singular characteristic of the first approximate solution v_0 .

Another way of seeing that a curve in whose neighbourhood the ordinary series approach breaks down must be a characteristic is to use curvilinear coordinates x, y such that the said curve is the line $x=0$. Now if v_0 has an algebraic or logarithmic singularity as $x \rightarrow 0$, then in its (linear) equation the coefficient of $d^2 v_0/dx^2$ must tend to zero if the other coefficients are to remain bounded. Hence in the equation for its characteristics (a homogeneous quadratic in dx and dy) the coefficient of dy^2 must tend to zero as $x \rightarrow 0$, showing that $x=0$ is a characteristic of the equation

for v_0 . (Purely elliptic equations, for which singularities are necessarily isolated points, are not treated in this paper. But the theory as here given would be applicable to an equation of mixed type like

$$(1-r^2)\partial^2 v/\partial r^2 + (r^{-1}-2r)\partial v/\partial r + r^{-2}\partial^2 v/\partial \theta^2 = 0,$$

known to aerodynamicists, for which most solutions in $r < 1$, where it is elliptic, are singular as $r=1$, a characteristic, is approached.)

The technique of §§ 2 and 3 is most easily applied if such a coordinate system be used, namely one with $x=0$ as the singular characteristic of v_0 . The exact equation will then in general (when solved for $\partial^2 v/\partial x^2$) be of the form

$$[p_0(x, y) + \alpha p_1(x, y, v, \partial v/\partial x, \partial v/\partial y, \partial^2 v/\partial x \partial y, \partial^2 v/\partial y^2) + \dots] \partial^2 v/\partial x^2 \\ = \text{terms in } \alpha, x, y, v, \partial v/\partial x, \partial v/\partial y, \partial^2 v/\partial x \partial y, \partial^2 v/\partial y^2 \text{ linear when } \alpha=0, \quad \dots \quad (37)$$

where $p_0(0, y)=0$. To render (37) more similar to equations previously discussed it may now be divided by p_0/x , a function regular on the curve $x=0$ if the zero of p_0 is simple. (In a practical problem the desirability of such a procedure would naturally be open to question.) The coefficient of $\partial^2 v/\partial x^2$ becomes $x+O(\alpha)$.

An alternative, more general, proof that for a linear equation coordinates x, y can always be found so that the coefficient of $\partial^2 v/\partial x^2$ is x , without the other coefficients being unbounded or all zero, is as follows: in characteristic coordinates ξ, η the equation is $\partial^2 v/\partial \xi \partial \eta = \text{terms independent of the second derivatives}$; if now the transformation $x=\xi\eta, y=\text{an independent function}$, be made, the coefficient of $\partial^2 v/\partial x^2$ becomes $(\partial x/\partial \xi)(\partial x/\partial \eta)=x$.

Treatment of the transformed equation in the form

$$\left. \begin{aligned} \partial v/\partial x &= u, \\ [x + \alpha p_1(x, y, v, u, \partial v/\partial y, \partial u/\partial y, \partial^2 v/\partial y^2) + \dots] \partial u/\partial x \\ &= \text{terms in } \alpha, x, y, v, u, \partial v/\partial y, \partial u/\partial y, \partial^2 v/\partial y^2 \text{ linear when } \alpha=0, \end{aligned} \right\}, \quad (38)$$

by expansions

$$\left. \begin{aligned} x &= z + \alpha x_1(z, y) + \alpha^2 x_2(z, y) + \dots, \\ u &= u_0(z, y) + \alpha u_1(z, y) + \dots, \\ v &= v_0(z, y) + \alpha v_1(z, y) + \dots, \end{aligned} \right\} \quad \dots \quad (39)$$

is now very similar indeed to that of equations like (24) and (26).

The reader should perhaps be advised to reduce the first two or three problems that he studies to the canonical form (38), so that in their solution he may argue by analogy with the solid body of fact established in §§ 2 and 3. After some experience he may prefer to apply the general technique of this paper more directly, using analogy more remotely, as is done in the following

Example.—The simplest wave equation for v , in characteristic coordinates, is $\partial u/\partial y=0$ where $\partial v/\partial x=u$. Its general solution is $u=u_0(x)$, and this may have a singularity on say $x=0$, where perhaps $u_0 \sim A x^{-a}$.

Now the wave equation is usually satisfied only approximately by a physical quantity, the exact equation being expressible in a form containing small non-linear terms, say $\partial u/\partial y = \alpha$ times a function of $x, y, u, v, \partial v/\partial y, \partial^2 v/\partial y^2$ and $\partial u/\partial x$. (The last derivative is the most important in the situation envisaged). The present technique will enable the real behaviour of u near $x=0$, where u_0 is singular, to be found. To fix the ideas, suppose that

$$\frac{\partial u}{\partial y} = \alpha \left(u + \frac{\partial y}{\partial v} \right) \frac{\partial u}{\partial x}, \quad \frac{\partial v}{\partial x} = u. \quad . \quad . \quad . \quad . \quad . \quad (40)$$

Substitution of the expansions (39) in (40) gives

$$\begin{aligned} (1 + \alpha x_{1z} + \alpha^2 x_{2z} + \dots)(u_{0y} + \alpha u_{1y} + \dots) - (\alpha x_{1y} + \alpha^2 x_{2y} + \dots) \\ \times (u_{0z} + \alpha u_{1z} + \dots) = (\alpha u_{0z} + \alpha^2 u_{1z} + \dots) \\ \times \left[u_0 + \alpha u_1 + \dots v_{0y} + \alpha v_{1y} + \dots - \frac{(\alpha x_{1y} + \dots)(v_{0z} + \alpha v_{1z} + \dots)}{1 + \alpha x_{1z} + \dots} \right], \end{aligned} \quad . \quad . \quad . \quad (41)$$

where suffixes denote partial derivatives, and $\partial/\partial x$ has been replaced by $x_z^{-1} \partial/\partial z$, and $(\partial/\partial y)_x$ by $(\partial/\partial y)_z - x_y x_z^{-1} \partial/\partial z$. The term independent of α gives $u_{0y} = 0$, whence $u_0 = u_0(z)$: the coefficient of α in (41) is

$$u_{1y} = u_0'(z)[x_{1y} + u_0 + v_{0y}]. \quad . \quad . \quad . \quad . \quad . \quad (42)$$

Hence if $u_0 \sim A z^{-q_0}$ ($q_0 > 0$), u_1 may be rendered innocuous (say $O(z^{-2q_0})$) by an x_1 satisfying $x_1 \sim -A y z^{-q_0}$ as $z \rightarrow 0$. As in §2 the radius of convergence of the expansions (39) may be found to be of order z^{q_0} as $z \rightarrow 0$, and as in §2 this means that, for small α , $x=0$ is an interior point of a region of validity of the expansions independent of α . The roughest approximation of value near $x=0$ is

$$u = u_0(z), \quad x = z - \alpha A y z^{-q_0}, \quad v = \int u_0(z) dz + F(y). \quad . \quad (43)$$

There is a line of branch-points (on which $\partial(x, y)/\partial(z, y) = 0$ and therefore the three second derivatives of v are infinite like the inverse square root of distance from the line) when

$$z = -(\alpha A y q_0)^{1/(1+q_0)}, \quad x = -(1 + q_0^{-1})(\alpha A y q_0)^{1/(1+q_0)}, \quad . \quad (44)$$

which is a real curve if $A y > 0$.

If alternatively in this example $q_0 \leq -1$, x_1 may be taken as $-y u_0(0) - v_0(0, y)$ to give a value of u_1 no more singular than u_0 . Finally if $u_0(z) = u_0(0) + A z^{-q_0} + O(z)$, with $-1 < q_0 < 0$, it would be sufficient to take $x_1 = -y(u_0(0) + A z^{-q_0}) - v_0(0, y)$.

§6. THE SPREAD OF A PROGRESSIVE WAVE AT INFINITY.

The nature of the non-uniformity of an approximate solution which will be considered in this section, and removed by an application of the general technique described in §1, is essentially different from what has gone before, and could be displayed only in a partial differential equation (of hyperbolic type).

Physically it concerns the behaviour of a progressive wave after it has travelled a distance large compared with its width. Assume that the hyperbolic equation of some wave motion is approximately linear and exactly quasi-linear (that is, linear in the second order derivatives). Then in the characteristic coordinates (x, y) of the linearized equation (which for wave propagation in the z -direction with constant velocity c would be $ct \pm z$) the exact equation is of the form

$$\frac{\partial^2 v}{\partial x \partial y} + F = A \frac{\partial^2 v}{\partial x^2} + B \frac{\partial^2 v}{\partial x \partial y} + C \frac{\partial^2 v}{\partial y^2} + D, \quad . . . \quad (45)$$

where F is linear (and A, B, C at least linear, and D at least of the second order), in $v, \partial v/\partial x$ and $\partial v/\partial y$, with coefficients functions of x and y .

Small wave motions satisfy approximately $\partial^2 v/\partial x \partial y + F = 0$; and if $F \rightarrow 0$ as $x \rightarrow \infty$ it might be inferred that in one class of solutions (roughly speaking, "progressive waves") v is propagated unchanged along characteristics $y = \text{constant}$, asymptotically as $x \rightarrow \infty$. This conclusion is erroneous, even when correct for the linearized equation. The non-linear terms, though small (and even perhaps decreasing, as will be true in a modification of the problem to be given directly), alter the wave motion completely after a sufficient time has elapsed.

More precisely suppose that, as $x \rightarrow \infty$,

$$F = \frac{\partial v}{\partial y} \left(\frac{n}{x} + O\left(\frac{1}{x^2}\right) \right) + \frac{\partial v}{\partial x} O\left(\frac{1}{x}\right) + v O\left(\frac{1}{x^2}\right), \quad . . . \quad (46)$$

where $n \geq 0$, and that in D similarly the coefficients are $O(x^{-1})$, while in A, B, C , the coefficients are $O(1)$. Then an asymptotic solution to the linearized equation, representing a progressive wave, is

$$v \sim \frac{v_0(y)}{x^n}, \quad \quad (47)$$

for which $x^n v$ is propagated unchanged (asymptotically) along lines $y = \text{constant}$. In fact a series solution for v in descending powers of x with (47) as largest term can be substituted in the linearized equation, and the coefficient of the highest power of x , in the resulting equation, namely x^{-n-1} , arises from $\partial^2 v/\partial x \partial y + nx^{-1} \partial v/\partial y$ and is zero. All the later terms can then be made zero by choice of the subsequent coefficients in the series for v , in terms of the arbitrary function v_0 .

But this conclusion will be invalidated by the presence of the non-linear terms in (45), at any rate of $C \partial^2 v/\partial y^2$ (the others are of smaller order than x^{-n-1} if $n > 0$). Now $C \partial^2 v/\partial y^2$ may be of order as large as x^{-2n} , so that if $n \leq 1$ it cannot be annulled by suitable choice of any terms smaller than (47) in v . (It may be remarked here that the equations of plane, cylindrical and spherical sound waves have $n = 0, \frac{1}{2}$ and 1 respectively and so all satisfy $0 \leq n \leq 1$.) This section will be concerned first with finding the asymptotic form of the *exact* wave motion as $x \rightarrow \infty$, and then with finding an approximate solution valid uniformly for all x , large or small.

It is convenient as in §5 to write equation (45) so that the crucial derivatives are of the first order, namely as

$$\frac{\partial u}{\partial x} + F = A \frac{\partial^2 v}{\partial x^2} + B \frac{\partial u}{\partial x} + C \frac{\partial u}{\partial y} + D, \quad \frac{\partial v}{\partial y} = u. \quad (48)$$

The linearized equation gives $u \sim v'_0(y)x^{-n} = u_0(y)x^{-n}$. Now at each point (x, y) with x large it must be *locally* true that x^nu is propagated unchanged along a line very closely of the form $y = \text{constant}$. More precisely, curves with $x^nu = \text{constant}$ satisfy $dy/dx = -(\partial u/\partial x + nu/x)/(\partial u/\partial y)$, which is small by (48) and (46). But from this fact it can only be concluded that in reality x^nu is propagated unchanged (for large x) along curves, with dy/dx everywhere small, but not necessarily with y approximately constant over a *long* stretch of values of x . This suggests that, as was stipulated as allowable in the statement of the general technique in §1, the "other" independent variable y , *not* the variable x whose tendency to infinity produces the non-uniformity, should be transformed; and that $u \sim u_0(z)x^{-n}$, with dy/dx small on the curves $z = \text{constant}$, may give a true picture of the exact behaviour.

It will now be found that by a suitable transformation $y = y(x, z)$, $x = x$, which is progressively determined with the various terms in a solution $u = u_0(z)x^{-n} + \dots$, the exact solution can be approximated for large x as closely as may be desired. But the nature of the variable z can be predicted in advance. For in (48) it is the term in $\partial u/\partial y$ in the right-hand side that makes a direct expansion $u = u_0(y)x^{-n} + \dots$ impossible. Hence the transformation will be effective if in the transformed equation no term in $\partial u/\partial z$ appears. This means that, expressed in terms of v , no term in $\partial^2 v/\partial z^2$ appears, in other words that the curves $z = \text{constant}$ are exact characteristics of the equation. Hence the following principle is deduced for the application of the technique of this paper to finding the possible asymptotic behaviour of a progressive wave: "If on a linearized theory u (the derivative of v across the characteristics) can be expanded in descending powers of x with coefficients constant on each approximate characteristic $y = \text{constant}$, seek a similar expansion with coefficients constant on each exact characteristic $z = \text{constant}$, finding the latter curves by a second similar expansion of y ."

Of course the expansions might still be possible if the curves $z = \text{constant}$ were not quite the exact characteristics, since a term $\partial u/\partial z$ can be tolerated in the equation for u if it is multiplied by a factor sufficiently small as $x \rightarrow \infty$: thus the possibility of the expansions alone does not determine z uniquely. Again, the condition that on $z = \text{constant}$ the characteristic equation

$$-dx dy = A dy^2 - B dx dy + C dx^2 \quad (49)$$

shall be satisfied does not determine z completely, unless the condition that u (which occurs in A, B, C) satisfies (48) be also used. But if the above principle is applied in its entirety the quantity z is uniquely determined for each possible motion, being on each characteristic the

constant term in the expansion of y in descending powers of x on that characteristic. In the solution for u and v , if no boundary conditions are applied, there appears one arbitrary function $u_0(z)$, and a singly infinite set of arbitrary constants, derived from integrating the second of equations (48), expanded in series. (This corresponds to the degree of arbitrariness expected in the solution to a second order partial differential equation.) The constants can be successively determined from the condition, till now rather vaguely expressed, that the wave be "progressive". This may mean that there is one "initial" characteristic on which $u=v=0$; this will be given by $u_0(z)=0$, and the constants must be determined so that for this value of z every coefficient in the series for u and v vanishes (of course on a characteristic the conditions on u and v are not independent). Alternatively in some problems the corresponding condition may be that the wave motion is bounded by a curve of discontinuity (a shock wave in gas dynamics), unknown a priori, on which a certain double boundary condition holds which suffices to determine both its position and the aforesaid system of arbitrary constants.

The root of (49) with dy/dx small is

$$\frac{dy}{dx} = \frac{1}{2A} (-1 + B + \sqrt{[(1-B)^2 - 4AC]}) \doteq -C. \quad . \quad . \quad (50)$$

Thus, to a first approximation, $y = z - \int C dx$, where the integration is along a curve $z = \text{constant}$. Since in general $C = O(x^{-n})$, this means that $y = z + O(x^{1-n})$, or $z + O(\log x)$ if $n=1$. Thus $y-z$ is infinite as $x \rightarrow \infty$: the spread of the characteristics is infinite, although the rate of spread is small and (for $n > 0$) decreasing. For $n=0$ they are roughly straight lines fanning out; for $n=\frac{1}{2}$ parabolæ $y = z + O(x^{\frac{1}{2}})$; for $n=1$ they spread out only at a slow logarithmic rate. (Even for $n < 1$ there may be terms in x^{-1} in the expansion of the right-hand side of (50), giving terms in $\log x$ in $y(x, z)$ and hence also logarithmic terms in the power series for u and v .)

The behaviour of v for large x is not like that of u . For

$$\frac{\partial v}{\partial z} = u \frac{\partial y}{\partial z} \sim u_0(z) x^{-n} \left(-\frac{\partial}{\partial z} \int C dx \right) = O(x^{1-2n}), \quad . \quad . \quad (51)$$

so that the series for v with coefficients functions of z begins with a term of order x^{1-2n} , greater than does that for $u = \partial v / \partial y$. This is because the distance dy between adjacent curves $z = \text{constant}$ increases like x^{1-n} as $x \rightarrow \infty$.

The theory given above would hold for *any* wave propagation, even with initially large disturbances, if $n > 0$; for then at least the disturbance becomes small (like x^{-n}) as $x \rightarrow \infty$. But, in cases where the initial disturbance is small, it can be described by a linearized theory—or, more exactly, u can be expanded in ascending powers of α (without a term independent of α), if α is some parameter estimating the size of the disturbance: the first term satisfies the linearized equation. This

series, for given α , would diverge when x is large enough, since the first coefficient, namely the linearized solution for u , is $O(x^{-n})$ as $x \rightarrow \infty$, while the second, which satisfies the linearized equation with a right-hand side including terms like $C\partial u/\partial y$ derived from the first term, would be $O(x^{1-2n})$; and the increase would continue.

But it can now be seen how a uniformly valid series solution (in powers of α) can be obtained. If a transformation

$$y = y(x, z) = z + \alpha y_1(x, z) + \alpha^2 y_2(x, z) + \dots, \quad (52)$$

is made, so that the curves $z = \text{constant}$ are exactly characteristics, that is, satisfy (49), and if also

$$u = \alpha u_0(x, z) + \alpha^2 u_1(x, z) + \dots, \quad (53)$$

is made to satisfy equation (48), then it is already known that no terms of order larger than $O(x^{-n})$ at infinity can occur in the series for u , or terms larger than $O(x^{1-n})$ in that for y , and therefore by the principle of §1 a uniform convergence in some range of α is to be expected.

A first approximation valid uniformly for unbounded x can now be obtained easily. By (52) and (49), $-\partial y_1/\partial x$ is simply equal to the linear terms in the function C , with u replaced by $u_0(x, z)$. (For $-\partial y/\partial x = A(\partial y/\partial x)^2 - B(\partial y/\partial x) + C = O(\alpha^3) + O(\alpha^2) + C$.) Hence if $C_1(x, z)$ signifies these terms, $y = z - \alpha \int C_1(x, z) dx + O(\alpha^2)$. But further, as was noticed in a similar situation in §2, to a first approximation $\int C_1(x, z) dx$ may be replaced by its asymptotic form as $x \rightarrow \infty$. For except when x is large, the term αy_1 in y alters the value of z by only a small quantity (and hence that of u by only a small multiple of itself) for given (x, y) ; and, when x is large, terms in αy_1 other than its first approximation as $x \rightarrow \infty$, cannot alter the value of $z(x, y)$ by more than the order of their ratio to the first approximation, which again is small.

Lastly it may be remarked that since $y = z$ for small x , the function $u_0(x, z)$ in (53) must be the function $u_0(x, y)$ which is the solution of the linearized equation describing the motion for bounded x : hence the following principle (the most useful conclusion of this section) for obtaining a uniformly valid approximation to a progressive wave motion in an unbounded domain is deduced:

If the linearized equation (in characteristic coordinates x, y) and boundary conditions give $u = \alpha u_0(x, y) \sim \alpha u_0(y)x^{-n}$ as $x \rightarrow \infty$, then take $u = \alpha u_0(x, z)$, where z is defined by $y = z - \alpha \int C^(x, z) dx$; and αC^* is the asymptotic form as $x \rightarrow \infty$ of the coefficient C of $\partial u/\partial y$ in the non-linear terms (to which in the exact equation of motion (48) the linear terms are equated), with $\alpha u_0(z)x^{-n}$ substituted for u therein.*

It is believed that the reader will find this principle more convincing expounded as above, representing the first terms in an infinite series expansion, than if it had been stated by itself with rougher arguments to support it. An example will now be given, chosen to limit the calculations required to be followed by the reader, rather than from any physical

theory. (An interesting and important example, concerning axisymmetrical supersonic flow, with $n=\frac{1}{2}$, has been worked out by Mr. Whitham, and will be published roughly simultaneously with this paper.)

Example.—Solve

$$\frac{\partial u}{\partial x} + \frac{n}{x+y}u = u \left(\frac{\partial^2 v}{\partial x^2} + \frac{\partial u}{\partial y} \right), \quad \frac{\partial v}{\partial y} = u, \quad (54)$$

with $u=v=0$ when $y=0$ and $u=\alpha U(y)y^{-n}$ on $x=0$, where α is small, $0 < n < 1$, and $U(0)=0$. The linearized solution is $u=\alpha U(y)(x+y)^{-n}$. Hence, by the principle italicized above, a uniformly valid approximation to order α is

$$u=\alpha U(z)(x+z)^{-n}, \quad y=z-\alpha U(z)x^{1-n}/(1-n). \quad . . (55)$$

To illustrate the whole theory a uniformly valid approximation to order α^2 is now obtained. In variables x, z , where $y=y(x, z)$, (54) becomes

$$u_x - \frac{y_x}{y_z}u_z + \frac{nu}{x+y} = u \left[\left(\frac{\partial}{\partial x} - \frac{y_x}{y_z} \frac{\partial}{\partial z} \right) (v_x - y_x u) + \frac{u_z}{y_z} \right], \quad v_z = uy_z. \quad . (56)$$

If $z=\text{constant}$ are characteristics, then

$$y_x = -uy_x^2 - u, \quad (57)$$

and (56) simplifies (losing its u_z terms) to

$$u_x + \frac{nu}{x+y} = u(v_{xx} - y_{xx}u - 2y_x u_x), \quad v_z = uy_z. \quad . . . (58)$$

Substitution of the series (52), (53) and a third, namely

$$v = \alpha v_0(x, z) + \alpha^2 v_1(x, z) + \dots,$$

in (57) and (58) gives firstly

$$u_0(x, z) = \frac{U(z)}{(x+z)^n}, \quad y_1(x, z) = -\frac{U(z)(x+z)^{1-n}}{1-n}, \quad v_0(x, z) = \int_0^z \frac{U(t)}{(x+t)^n} dt, \quad . . . (59)$$

and secondly

$$\frac{\partial u_1}{\partial x} + \frac{nu_1}{(x+z)} = \frac{ny_1 u_0}{(x+z)^2} + u_0 \frac{\partial^2 v_0}{\partial x^2}, \quad \frac{\partial y_2}{\partial x} = -u_1. \quad . . . (60)$$

From (59) and (60),

$$\frac{\partial}{\partial x} [(x+z)^n u_1] = -\frac{nU^2(z)}{(1-n)(x+z)^{1+n}} + U(z) \int_0^z \frac{n(n+1)U(t)}{(x+t)^{n+2}} dt. \quad . . . (61)$$

But, on $x=0$, $u=\alpha U(y)y^{-n}=\alpha u_0(0, y)$. Hence

$$\alpha u_0(0, y + \alpha y_1 + \dots) + \alpha^2 u_1(0, z) + \dots = \alpha u_0(0, y)$$

or

$$u_1(0, z) = -[du_0(0, z)/dz]y_1(0, z) = \left[\frac{d}{dz} (U(z)z^{-n}) \right] \frac{U(z)z^{1-n}}{1-n}.$$

Hence by (61)

$$u_1 = \frac{U^2(z)}{1-n} \left[\frac{1}{(x+z)^{2n}} - \frac{1}{z^n(x+z)^n} \right] + \frac{U(z)z^{1-n}}{(1-n)(x+z)^n} \frac{d}{dz} (U(z)z^{-n}) \\ + \frac{nU(z)}{(x+z)^n} \int_0^z \left(\frac{1}{t^{n+1}} - \frac{1}{(x+t)^{n+1}} \right) U(t) dt. \quad (62)$$

For large x , $u_1(x, z) \sim F(z)x^{-n}$, where

$$F(z) = -\frac{U^2(z)}{(1-n)z^n} + \frac{U(z)z^{1-n}}{1-n} \frac{d}{dz} (U(z)z^{-n}) + nU(z) \int_0^z \frac{U(t)}{t^{n+1}} dt. \quad (63)$$

Thus the asymptotic behaviour of u for large x is

$$u \sim \frac{\alpha U(z) + \alpha^2 F(z) + \dots}{x^n} \text{ as } x \rightarrow \infty. \quad (64)$$

(It is of course the whole numerator of (64) that is lumped together as " $u_0(z)$ " in the theory of the asymptotic behaviour of a progressive wave which occupied the first half of this section.) Lastly, by the second of equations (60),

$$y_2 \sim -\frac{F(z)x^{1-n}}{1-n} \text{ as } x \rightarrow \infty. \quad (65)$$

To obtain an approximation to order α^2 , valid uniformly, it is necessary to include from y_2 only this asymptotic form, putting

$$\left. \begin{aligned} u &= \alpha u_0(x, z) + \alpha^2 u_1(x, z), \\ y &= z + \alpha y_1(x, z) - \alpha^2 F(z)x^{1-n}/(1-n). \end{aligned} \right\} \quad (66)$$

For the change in z for given (x, y) , produced by including a term $\alpha^2 y_2$ in addition to $y = z + \alpha y_1$, is of order $\alpha^2 y_2 / (\partial y / \partial z) = \alpha^2 y_2 / (1 + O(\alpha x^{1-n}))$. This can be of order α for large x , causing u to be altered by a factor $1 + O(\alpha)$; but the effect of other terms than (65) in y_2 on $z(x, y)$ is uniformly $o(\alpha)$.

REFERENCES.

- LIGHTHILL, 1948, *Quart. J. Mech. & App. Math.*, **1**, 309-318.
POINCARÉ, 1892, *Les méthodes nouvelles de la mécanique céleste*, **1**, Ch. 3 (Paris).

CX. *The Shock Strength in Supersonic "Conical Fields."*

By M. J. LIGHTHILL *.

[Received August 22, 1949.]

SUMMARY.

In the linearized theory of supersonic "conical fields," there is a singularity on the Mach cone from the apex. In this paper, using the exact equations of motion, and a general method of rendering approximate solutions uniform, which is described elsewhere, but is here fully explained as far as is needed in the present application, the true behaviour of the flow near the Mach cone is deduced. Near to it is a shock; the most interesting conclusion is that the strength of the shock can be inferred, to a first approximation, from the linearized solution alone (by formulæ derived below from the exact equation of motion). The general flow, with disturbing surfaces outside as well as inside the Mach cone, is treated in § 5. In the simpler special case, treated in § 4, when the disturbance on linear theory is confined within the Mach cone, the results can be stated simply. Let the intersection of the cone with a plane perpendicular to the stream be the circle $r=1$ in polar coordinates in that plane; then the pressure on linearized theory has a singular behaviour as $r \rightarrow 1$ of the type $(p-p_1)/\rho_1 U^2 \sim A(\theta)\sqrt{1-r}$, and it is shown below that the shock (a conical surface slightly outside the Mach cone) has strength which varies with θ like $\frac{3}{4}\gamma(\gamma+1)M^6(M^2-1)^{-1}A^2(\theta)$, to a first approximation, at points where $A(\theta) > 0$; but which is at most of order the cube of the disturbance velocities at points where $A(\theta) \leq 0$, when the most noticeable phenomenon is a limited region of rapid expansion, at a finite rate independent of the size of the disturbance. Examples are given in § 6 from a number of problems whose linearized solution is known. In the supersonic flow past any conical surface all of whose generators make a small angle with the main stream, the shock is found to be of a uniform strength to a first approximation, which depends only on the solid angle of the cone; two examples from supersonic wing theory are also given.

The work explains in detail why linear theory gives, as all writers on the subject have assumed, a correct first approximation to the flow field within the Mach cone, provided that the boundary condition of continuous velocity across the cone be applied, even though the linear solution is essentially the first term in a series, which diverges near the Mach cone, and although a shock discontinuity is really present in its neighbourhood.

§1. INTRODUCTION.

VALUABLE information on the three-dimensional steady supersonic flow past wings has been obtained from the study of "conical fields," which Busemann (1943) initiated. In these the velocity vector is constant along every line through some point (the "apex" of the conical field), and therefore in spherical polars with the apex as origin it depends only on the angular variables. Such a field will occur in any flow where the boundary conditions define naturally no linear dimension. Indeed, even where this is not so, part of the field may be conical, owing to

* Communicated by the Author.

the property of supersonic flow that disturbances cannot be propagated upstream; thus if the boundary conditions differ only in a certain region from conditions which define no linear dimension, then the velocity field will still be conical upstream of this region (or, more precisely, outside its "range of influence").

Most of the study of conical fields has been based on the well-known linearized equation of steady irrotational gas flow. The resulting theory is mathematically elegant as well as practically valuable, and recently a full account of it has been given by Goldstein and Ward (1949), together with references to the earlier literature. It enables the pressure distribution over any flat triangular wing at small incidence to be found, provided that the trailing edge is at an angle greater than the Mach angle to the main stream; for if this is so the range of influence of the trailing edge, where the conical field is "cut off", does not include the wing itself. Quadrilateral wings can also be treated; thus in flow past a rectangular wing there are two cone fields, one from each end of the leading edge; even if the aspect ratio is small enough for these to interact, an additive treatment of the interaction is possible on a linear theory; only if one of the two Mach cones intersects the opposite wing tip does the method break down.

In the linearized theory, let the undisturbed velocity be in the z -direction, with Mach number M . The conical field condition, with the apex as origin of cartesian coordinates, can be expressed by saying that each component of the velocity depends only on the variables x/z and y/z . Putting $\beta x/z = r \cos \theta$, and $\beta y/z = r \sin \theta$, where $\beta = \sqrt{M^2 - 1}$, so that the circle $r=1$ corresponds to the Mach cone from the apex, and expressing that the velocity potential satisfies the linearized equation $\nabla^2 \phi = M^2 \partial^2 \phi / \partial z^2$, it is deduced that each component of velocity satisfies a certain equation

$$(1-r^2) \frac{\partial^2 u}{\partial r^2} + \left(2r + \frac{1}{r}\right) \frac{\partial u}{\partial r} + \frac{1}{r^2} \frac{\partial^2 u}{\partial \theta^2} = 0, \quad . \quad . \quad . \quad (1)$$

which is elliptic inside the Mach cone (where $r < 1$) and hyperbolic outside it (where $r > 1$).

By a transformation it is shown (Goldstein and Ward 1949) that solutions of (1) in $r < 1$ are solutions of Laplace's equation in the variables $\text{sech}^{-1} r$ and θ , and that solutions of (1) in $r > 1$ are solutions of the wave equation in the variables $\sec^{-1} r$ and θ . These facts, with certain relations between the relevant solutions for the three components of velocity, enable the completion of problems in terms of the relevant boundary conditions.

The shortcoming of the linear theory is that it introduces a singularity on the Mach cone from the apex, which cannot describe correctly the real behaviour of the gas. For since $\text{sech}^{-1} r \sim \sqrt{2(1-r)}$ and $\sec^{-1} r \sim \sqrt{2(r-1)}$ as $r \rightarrow 1$, solutions of (1) in general will have square root singularities of these types as $r \rightarrow 1$ from below and above respectively (though it must be remarked that in many practical problems the velocity is constant as $r \rightarrow 1$ from above for given θ , so that only the singularity as $r \rightarrow 1$ from below actually arises).

It is the purpose of this paper to find, using the full non-linear equations, what really happens near $r=1$. One has experimental evidence for the presence of a shock, which it will be found is borne out by the theory. An interesting aspect of the discoveries is that the strength of the shock can always be deduced (to a first approximation) from the nature of the singularity of the linearized equation alone, although the exact equations must be used to deduce the formula by which this is done (which indeed involves γ , the adiabatic index, absent from the linearized equation), and although the strength is of the second order in the disturbance velocities. As a result, in all the problems which have so far been solved by the linearized theory of conical fields, the formulæ derived below enable one to calculate immediately how the shock strength varies round the Mach cone $r=1$.

The theory used still assumes the flow field to have constant entropy, and hence zero vorticity; but this is permissible provided that no deductions are drawn relating to quantities of order as small as the cube of the strengths of any shocks which occur. On this assumption, the exact equations of motion in the (r, θ) plane, and the form of the conditions across a shock, are obtained in § 2. Next a general method of rendering uniformly valid an approximate solution to a partial differential equation under given boundary conditions, which is described in its general form in another paper (1949) by the author, is applied to the present problem. The nature of the method in this application is discussed fully in § 3. The problem is solved in § 4 in the case when there is no disturbing surface outside the Mach cone $r=1$. It is found that if the pressure coefficient $(p-p_1)/\rho_1 U^2$ behaves on the linearized theory like $A(\theta)\sqrt{1-r}$ as $r \rightarrow 1$, then the strength of the nearly conical shock (*i.e.* the relative pressure change across it) varies with θ asymptotically like $\frac{3}{4}\gamma(\gamma+1)M^6\beta^{-2}A^2(\theta)$, provided that $A(\theta) > 0$. (This agrees with the value for axial flow past a circular cone of small semi-angle α , found previously by the author (1948), where $A(\theta) = \alpha^2\sqrt{2}$). For values of θ with $A(\theta) < 0$ the shock strength is zero to this order; behind the shock however, there is a finite (non-small) pressure gradient $\partial p/\partial r = \rho_1 U^2 \beta^2/(\gamma+1)M^4$, and a gradient of this order is maintained for a distance of order the square of the disturbances. In § 5 the considerably more complicated general problem, with disturbing surfaces outside the Mach cone, is solved. In § 6 some conclusions are drawn in particular cases.

§2. EXACT EQUATIONS OF A STEADY IRROTATIONAL CONICAL FIELD.

In cylindrical coordinates (ρ, θ, z) the equation for the potential in the flow of a gas with constant adiabatic index γ at constant entropy is

$$\alpha^2(\phi_{\rho\rho} + \rho^{-1}\phi_{\rho} + \rho^{-2}\phi_{\theta\theta} + \phi_{zz}) = \frac{1}{2} \left(\phi_{\rho} \frac{\partial}{\partial \rho} + \rho^{-1}\phi_{\theta} \frac{\partial}{\partial \theta} + \phi_z \frac{\partial}{\partial z} \right) (\phi_{\rho}^2 + \rho^{-2}\phi_{\theta}^2 + \phi_z^2),$$

. (2)

where the suffix derivative notation has been used, and a , the local speed of sound, is given by

$$a^2 = a_1^2 + \frac{1}{2}(\gamma - 1)(U^2 - \phi_e^2 - \rho^{-2}\phi_\theta^2 - \phi_z^2), \quad (3)$$

where a_1 is the velocity of sound when the fluid speed is U . U will be taken as the velocity of the undisturbed stream in §4, though in §5 it will be convenient to give it temporarily a slightly different meaning.

Now if $\phi = Uz(1 + f(r, \theta))$, where $r = \beta\rho/z$ is the same r as was introduced in §1, a conical field will be present, since the velocity components $\phi_z = U(1 + f - rf_r)$, $\phi_e = \beta U f_r$, and $\rho^{-1}\phi_\theta = \beta U r^{-1}f_\theta$ then depend only on r and θ . The equation of motion (2), on multiplication by z/U^2 , becomes

$$\begin{aligned} [M^{-2} - (\gamma - 1)(f - rf_r + \frac{1}{2}(f - rf_r)^2 + \frac{1}{2}\beta^2 f_r^2 + \frac{1}{2}\beta^2 r^{-2} f_\theta^2)](\beta^2 f_{rr} + \beta^2 r^{-1} f_r \\ + \beta^2 r^{-2} f_{\theta\theta} + r^2 f_{rr}) = \frac{1}{2} \left[\beta^2 f_r \frac{\partial}{\partial r} + \beta^2 r^{-2} f_\theta \frac{\partial}{\partial \theta} - (1 + f - rf_r)r \frac{\partial}{\partial r} \right] \\ \times [\beta^2 f_r^2 + \beta^2 r^{-2} f_\theta^2 + (1 + f - rf_r)^2]. \quad (4) \end{aligned}$$

Equation (4) contains terms linear, quadratic and cubic in f and its derivatives. Many of the cubic ones will not be needed in what follows; in fact only those involving f_{rr} are required. Let \mathbf{C} stand for cubic terms in (4) which do not contain f_{rr} . Then the right-hand side of (4) can be written

$$r^2(1 + f - rf_r)^2 f_{rr} - 2\beta^2 r(1 + f - rf_r)f_r f_{rr} + \beta^4 f_r^2 f_{rr} + 2\beta^2(r^{-2}f_\theta^2 - r^{-1}f_{r\theta})f_\theta + \mathbf{C}, \quad (5)$$

and (4) as a whole, multiplied by $M^2\beta^{-2}$, becomes

$$\begin{aligned} [1 - r^2 - (\gamma + 1)M^2\beta^{-2}r^2(f - rf_r + \frac{1}{2}(f - rf_r)^2) - (\gamma - 1)M^2\{f - rf_r + \frac{1}{2}(f - rf_r)^2 \\ + \frac{1}{2}(r^2 + \beta^2)(f_r^2 + r^{-2}f_\theta^2)\} + 2M^2rf_r(1 + f - rf_r) - M^2\beta^2 f_r^2]f_{rr} \\ + [1 - (\gamma - 1)M^2(f - rf_r)](r^{-1}f_r + r^{-2}f_{\theta\theta}) - 2M^2(r^{-2}f_\theta^2 - r^{-1}f_{r\theta})f_\theta + \mathbf{C} = 0. \quad (6) \end{aligned}$$

The linear terms in (6), namely $(1 - r^2)f_{rr} + r^{-1}f_r + r^{-2}f_{\theta\theta} = 0$, are not exactly as in (1), since it is only the velocity components $\phi_z = U(1 + f - rf_r)$, $\phi_x = \beta U(f_r \cos \theta - r^{-1}f_\theta \sin \theta)$ and $\phi_y = \beta U(f_r \sin \theta + r^{-1}f_\theta \cos \theta)$, which satisfy (1) on a linear theory. Equation (1) is simpler to solve than the linear equation for f . But if ϕ_x , ϕ_y , and ϕ_z are known from solving (1) then f_r and f are deduced from the equations

$$f_r = \frac{1}{\beta U}(\phi_x \cos \theta + \phi_y \sin \theta), \quad f = rf_r + \frac{1}{U}\phi_z - 1; \quad . . . (7)$$

it is easiest to work with f in what follows.

The conditions at a stationary shock were given in various forms by Rankine and Hugoniot independently; a simple form is

$$(i) \mathbf{q}_t \text{ continuous} \quad (ii) \Delta q_n = \frac{2}{\gamma + 1} \left(\frac{a^2}{q_n} - q_n \right); \quad . . (8)$$

where \mathbf{q}_t represents in magnitude and direction the velocity tangential

to the shock surface, q_n is the component normal to the shock and Δq_n is the change in q_n on passing through the shock in the direction of the stream. Condition (ii) is true if q_n and a are both measured directly ahead of the shock (and true with its sign changed if both are measured directly behind it).

If, as in this paper, vorticity created by the shock is neglected, so that a potential ϕ exists, then condition (i) is equivalent to the continuity of ϕ , and hence also of $f(r, \theta)$, at the curve in the (r, θ) plane which represents the conical (in the general sense) shock surface. If this surface is $r = \eta(\theta)$, in other terms $-\beta\rho + z\eta(\theta) = 0$, then a vector normal to it has radial, transverse and axial components proportional to $-\beta, z\rho^{-1}\eta'(\theta)$ and $\eta(\theta)$. Hence

$$q_n = \frac{-\beta\phi_\theta + z\rho^{-2}\eta'(\theta)\phi_\theta + \eta(\theta)\phi_z}{\{\beta^2 + z^2\rho^{-2}\eta'^2(\theta) + \eta^2(\theta)\}^{\frac{1}{2}}} = U \frac{\eta(1+f-rf_r) - \beta^2 f_r + \beta^2 r^{-2}\eta' f_\theta}{\{\beta^2 + \eta^2 + \beta^2 r^{-2}\eta'^2\}^{\frac{1}{2}}}, \quad (9)$$

which with (3) enables the right-hand side of (8) (ii) to be calculated. To obtain the left-hand side observe that $\Delta f = 0$ on $r = \eta(\theta)$, and hence (differentiating along the curve) that $\Delta(f_\theta + \eta'(\theta)f_r) = 0$. From these it follows that

$$\Delta q_n = U \frac{-\eta r \Delta f_r - \beta^2 \Delta f_r - \beta^2 r^{-2} \eta'^2 \Delta f_r}{\{\beta^2 + \eta^2 + \beta^2 r^{-2} \eta'^2\}^{\frac{1}{2}}} = -U \{\beta^2 + \eta^2 + \beta^2 r^{-2} \eta'^2\}^{\frac{1}{2}} \Delta f_r. \quad (10)$$

For some purposes a first order approximation to condition (8) (ii) is adequate, where terms quadratic in $\eta - 1, f$, or the derivatives of either are neglected. Then it becomes

$$\begin{aligned} \Delta f_r &= -\frac{2}{\gamma+1} \left\{ \frac{a^2}{U^2(\eta+f-M^2 f_r)} - \frac{\eta+f-M^2 f_r}{\beta^2 + \eta^2} \right\} + Q \\ &= -\frac{2}{\gamma+1} [(M^{-2} - (\gamma-1)(f-f_r))(1 - (\eta-1) - f + M^2 f_r) \\ &\quad - (1 + (\eta-1) + f - M^2 f_r)M^{-2}(1 - 2M^{-2}(\eta-1))] + Q, \quad (11) \end{aligned}$$

where Q stands for terms at least quadratic in $f, \eta - 1$ and their derivatives. Simplifying (11), the conditions (8) at the shock $r = \eta(\theta)$ become

(i) f continuous,

$$(ii) \Delta f_r = \frac{2}{\gamma+1} [(\gamma-1)(f-f_r) + 2M^{-2}f - 2f_r + 2M^{-4}\beta^2(\eta-1)] + Q. \quad (12)$$

§3. METHOD OF IMPROVING THE LINEAR APPROXIMATION NEAR $r=1$.

In this section the method to be employed in §§ 4 and 5 for solving the problem of the real behaviour of conical fields near $r=1$ will be described and explained. The reader will find, in the general account of the method (Lighthill 1949), how it was discovered, by the study of analogous, but very much simpler, first order ordinary differential equations; however, the author hopes that the present account of its application to equation (6) will be found complete in itself.

The investigation will pick out two solutions, each possessing repeated derivatives, one in a region approximately (but not exactly) $r > 1$, and one in a region approximately (but not exactly) $r < 1$; the regions overlap. The true solution will consist of the former for $r > \eta(\theta)$, and the latter for $r < \eta(\theta)$, where $r = \eta(\theta)$ must either be a shock (where the shock conditions must be satisfied), or else a characteristic (where the velocity, but not necessarily its derivatives, must be continuous); the latter may be regarded as a "shock of zero strength". It should be emphasized that $r = \eta(\theta)$ will not normally be the boundary of the region of regularity of either solution; it is merely the curve at which the change from the one to the other occurs.

It will now be described how each of these solutions can be found approximately (by the same method). An expansion $f = f_1 + f_2 + f_3 + \dots$ will be made, where f_1 represents the linearized value of f , f_2 consists of terms of quadratic order in f_1 and its derivatives, f_3 of terms of cubic order, and so on. It will be convenient to speak of f_n as "of dimension n ", which will mean essentially "of order of the n th power of the disturbance velocities introduced by the boundary conditions". When such a series is substituted in an equation, the terms of each dimension will be equated separately, under the convention that the dimension of a product is the sum of the dimensions of its separate terms. But this expansion will *not* be substituted directly in equation (6).

For, as was shown in §1, the linearized velocity components are regular functions of $(r-1)^{\frac{1}{2}}$ and θ near $r=1$ *. Hence by (7) the linearized value of f_r , namely f_{1r} , is a regular function of $(r-1)^{\frac{1}{2}}$ and θ . Thus f_{1rr} is in general infinite like $(r-1)^{-\frac{1}{2}}$ as $r \rightarrow 1$; though the other derivatives of f_1 of the first and second orders are finite, and though, in f_1 itself, the largest term not regular in r as $r \rightarrow 1$ is of order $(r-1)^{\frac{1}{2}}$ (by integration of f_{1r}). Now the equation for f_2 , if the expansion $f = f_1 + f_2 + \dots$ were substituted directly in (6), would be

$$(1-r^2)f_{2rr} + \frac{1}{r}f_{2r} + \frac{1}{r^2}f_{2\theta\theta} = -[\text{quadratic terms in (6) with } f \text{ replaced by } f_1]. \quad \dots (13)$$

These quadratic terms include the product of f_{1rr} with a linear expression in f_1 , namely $-(\gamma+1)M^2\beta^2r^{-2}(f_1 - rf_{1r}) - (\gamma-1)M^2(f_1 - rf_{1r}) + 2M^2rf_{1r}$. If (as is true "in general") this linear expression does not vanish identically when $r=1$, then the quadratic terms in (13) are infinite like f_{1rr} , that is like $(r-1)^{-\frac{1}{2}}$, as $r \rightarrow 1$, and hence f_{2r} itself must also be infinite of this order. When r is *not* near 1, f_{2r} is of order the square of the disturbance velocities, and so is small compared with f_{1r} . But as $r \rightarrow 1$ this order of magnitude, multiplied by $(r-1)^{-\frac{1}{2}}$, ultimately exceeds that of f_{1r} ; and so there is a region near $r=1$ where the second term in the approximation by series exceeds the first. The first alone has become a worthless

* When $r > 1$, and similarly of $(1-r)^{\frac{1}{2}}$ and θ when $r < 1$. For the time being, this gloss will not be repeated in what follows, and should be understood.

approximation; but so is the sum of first and second, for the third is similarly found to have become larger still, and indeed the whole series diverges in this region, where it gives no information about the true f whatsoever.

Furthermore, even if the coefficient of f_{1rr} in (13) did vanish identically, so that the f_{2r} , derived from solving (13) under the appropriate boundary conditions, was not infinite at $r=1$ but (like f_{1r}) a regular function of $(r-1)^{\frac{1}{2}}$ and θ , then nevertheless f_3 would satisfy an equation like (13), with the quadratic terms replaced by cubic terms (some involving f_1 alone and some f_1 and f_2 together); these would include terms in f_{1rr} and in f_{2rr} ; and unless the coefficients happened to be identically zero when $r=1$ (actually that of f_{2rr} is the coefficient of f_{1rr} in the equation for f_2 , which has already been assumed to vanish on $r=1$, so that it is only a question of whether the other vanishes), it would follow that f_{3r} is infinite at $r=1$, with consequent divergence (and also uselessness) of the whole series in some region near $r=1$.

The whole process would only be valid if it happened that the terms of *each* dimension in the coefficient of f_{rr} in (6) vanished separately when $r=1$ (when f is replaced by the series $f_1+f_2+\dots$ and f_n is treated as "of dimension n "); an infinity of conditions that in fact cannot be simultaneously satisfied except in trivial cases. But these conditions point to what the correct method must be, namely:—*

Make a transformation to new variables R and θ (this R corresponds to " z " in the general theory (Lighthill 1949)), so chosen that the transformed equation is like equation (6), but is such that the conditions described in the last paragraph hold (with R for r). Clearly the exact form of such a transformation will have to be determined progressively as successive terms in the series for f are found. This indicates a double expansion †

$$\left. \begin{aligned} r &= R + r(\theta) = R + r_1(\theta) + r_2(\theta) + r_3(\theta) + \dots, \\ f &= f_1(R, \theta) + f_2(R, \theta) + f_3(R, \theta) + \dots, \end{aligned} \right\} \quad (14)$$

where r_k , like f_k , is of dimension k .

When the series expansions (14) are substituted in (6), a new equation with independent variables R and θ will result. The coefficient of f_{RR} in this, or rather of $f_{1RR} + f_{2RR} + \dots$, must now be made to vanish at $R=1$, to every dimension, by the following process:

- (i) Determine f_1 (which will be the same function of R, θ as it was previously of r, θ —since the terms $r_1(\theta)$ etc., in the transformation affect equation (6) only to dimensions higher than the first).

* To help the reader, certain complications which do not affect the application of the following method in the later sections, though they do affect the demonstration of its validity, have been omitted from the text and relegated to a series of footnotes. These complications, though to be expected from the general theory (Lighthill 1949), might confuse those who encounter it first in this paper, who are therefore advised to ignore the footnotes.

† But it will be found that r_k must be a function of R as well as θ for $k \geq 2$.

- (ii) Choose $r_1(\theta)$ to make the terms of dimension 1 in the coefficient of f_{RR} vanish at $R=1$.
- (iii) Determine f_2 .
- (iv) Choose $r_2(\theta)$ to make the terms of dimension 2 in the coefficient of f_{RR} vanish at $R=1$.
- (v) etc. etc.

At each stage the equation for f_k will be like (13), with R for r , but with the term in square brackets replaced by a regular function of $(R-1)^{\frac{1}{2}}$ and θ , since any terms f_{jRR} in this (with $j < k$) are multiplied by a coefficient vanishing at $R=1$. Hence each f_{kR} will itself be a regular function of $(R-1)^{\frac{1}{2}}$ and θ , and the progressive increase with k in order of magnitude near $R=1$ will have been avoided, giving convergence just as much near $R=1$ as elsewhere, and enabling the problem of this paper to be solved*.

For the purposes of this paper only the four steps (i), (ii), (iii) and (iv), written out explicitly above, will be used; and in fact (iii) will be curtailed to read "determine a relation between certain derivatives of f_2 at $R=1$ ", this being all that is necessary to enable (iv) to be completed. But the possibility of indefinite continuation of the process, indicated in the last paragraph, is essential to its justification.

As was stated at the beginning of this section, two solutions will be found, that is to say the process will be carried through to give two sets of values for the functions in (14). One set will hold for $R > 1$, the other for $R < 1$ (indeed extension through $R=1$ is impossible, owing to terms like $(R-1)^{\frac{1}{2}}$, $(1-R)^{\frac{1}{2}}$ respectively in the two solutions). *But these two regions overlap*, because R has a different meaning (for given r, θ) in the

* Here a difficulty has been glossed over. For each term of order $(R-1)^n$ on the right-hand side of (13) a term of the same order in f_{2R} will correspond, except when $n=\frac{1}{2}$, as the reader will verify. In this case the corresponding term in f_{2R} contains a $(R-1)^{\frac{1}{2}} \log(R-1)$. Therefore such a term must in general occur in f_{2R} in addition to terms regular in $(R-1)^{\frac{1}{2}}$ and θ . Hence it is necessary to modify r_2 from the value given by the process in the text, lest a term $\log(R-1)$ should appear in f_3 , leading to terms of higher order later on. In fact r_2 must not be a mere function of θ , chosen to make the terms of dimension 2 in the coefficient of f_{RR} vanish at $R=1$, but it must be made to depend on R as well, so that these terms not only vanish but are $O(R-1)$ as $R \rightarrow 1$. This will not affect the value of r_2 at $R=1$, which is fixed by the requirement that the terms should vanish there, and is all that will be needed in what follows; it merely adds thereto a term in $(R-1)^{\frac{1}{2}}$. Hence the theory given in the text is adequate for what follows. The nature, as $R \rightarrow 1$, of the further terms in the expansions (14) is fully indicated by the general theory (Lighthill 1949); they each remain bounded, which indicates convergence, although terms involving a product of $(R-1)^{\frac{1}{2}}$ with higher and higher powers of $\log(R-1)$ occur in both expansions. This may mean (Lighthill 1949) that divergence exists when R is exponentially close to 1, so that $\log(R-1)$ is comparable with the reciprocals of the disturbance velocities. However even if this is so the results of this paper are unaffected, as it will not be needed to use the theory with R nearly so close to unity as this requires.

two regions ; and it is in the region of overlap that a shock (possibly of zero strength) will be drawn, at which the appropriate discontinuity conditions will have to be satisfied.

§4. DETERMINATION OF THE SHOCK WHEN NO DISTURBING SURFACE IS OUTSIDE THE MACH CONE.

When the solid boundary of the field lies entirely inside the Mach cone $r=1$, then on a linear theory the main stream is undisturbed for $r>1$. It will be found in this case that the exact equations can be satisfied if it be assumed that outside the shock (which lies near to $r=1$) the stream is undisturbed.

This is much simpler than the general case, since one of the two solutions mentioned at the end of the last section, namely the one for $R>1$, is very simple, being in fact $r=R$, $f\equiv 0$. Only the other requires detailed investigation, and for it the steps (i), (ii), (iii) and (iv) of the process of § 3 must be gone through.

As a result of stage (i) the linearised solution $f_1(R, \theta)$ will be obtained, which will be assumed known in what follows (having been obtained by methods fully discussed by Goldstein and Ward (1949) ; we have merely to substitute R for r). The boundary conditions $f_1=f_{1R}=0$ on $R=1$ (corresponding to zero disturbance velocity on the Mach cone) will have been applied in obtaining this solution. Therefore the asymptotic behaviour of f_{1R} as $R\rightarrow 1$ must be of the form

$$f_{1R}\sim A(\theta)\sqrt{1-R}; \quad . \quad . \quad . \quad . \quad . \quad (15)$$

and indeed if the linearized solution is given in the $(\text{sech}^{-1} R, \theta)$ plane, then $A(\theta)$ is simply $-2^{-\frac{1}{2}}(\partial f_{1R}/\partial \text{sech}^{-1} R)_{R=1}$. (In (15) the positive square root of $1-R$ must be understood). This $A(\theta)$ agrees with the definition in §1, since on linear theory

$$(p-p_1)/\rho_1 U^2=1-U^{-1}\phi_z=r f_r-f\sim A(\theta)\sqrt{1-R} \text{ as } R\rightarrow 1.$$

Now apply stage (ii) ; substitute the expansions (14) in (6) and express the condition that on $R=1$ the terms of dimension 1 in the coefficient of f_{RR} vanish. One must observe that in the transformation implied in (14) $\partial/\partial r$ will become $\partial/\partial R$ but $\partial/\partial \theta$ (keeping r constant) will become $\partial/\partial \theta - r'(\theta)\partial/\partial R$. Thus terms in $f_{r\theta}$ and $f_{\theta\theta}$ in (6) may produce terms in f_{RR} in the transformed equation. However it is seen that these are

$$[1-(\gamma-1)M^2(f-rf_R)]r^{-2}r'^2f_{RR}+2M^2r^{-1}(f_\theta-r'f_R)(-r')f_{RR}, \quad . \quad (16)$$

plus terms of dimension at least 4 ; and there are no terms of dimension 1 in the coefficient of f_{RR} in (16). Hence condition (ii) gives simply that

$$-2Rr_1-(\gamma+1)M^2\beta^{-2}R^2(f_1-Rf_{1R})-(\gamma-1)M^2(f_1-Rf_{1R})+2M^2Rf_{1R} \quad . \quad (17)$$

should vanish on $R=1$. Since actually $f_1=f_{1R}=0$ on $R=1$ this means that $r_1(\theta)=0$ identically.

Now, in stage (iii), simply write down the equation for f_2 with $R=1$ (which will be sufficient). This includes a term in f_{1RR} , which by (15) is asymptotically $-\frac{1}{2}A(\theta)(1-R)^{-\frac{1}{2}}$ as $R \rightarrow 1$. In the coefficient of f_{1RR} , which has been made to vanish at $R=1$, only the terms which are like $(1-R)^{\frac{1}{2}}$ as $R \rightarrow 1$ will contribute anything to the limit of the product. These are the terms in f_{1R} (since f_1 itself has no terms in $(1-R)^{\frac{1}{2}}$). Thus the limit of the product is

$$\lim_{R \rightarrow 1} [(\gamma+1)M^2\beta^{-2}R^3f_{1R} + (\gamma+1)M^2Rf_{1R}]f_{1RR} = -\frac{1}{2}(\gamma+1)M^4\beta^{-2}A^2(\theta). \quad (18)$$

All the other terms in f_1 in the equation for f_2 vanish as $R \rightarrow 1$, since $f_1=f_{1R}=f_{1\theta}=0$ on $R=1$, and it is deduced that on $R=1$,

$$f_{2R} + f_{2\theta\theta} = \frac{1}{2}(\gamma+1)M^4\beta^{-2}A^2(\theta). \quad (19)$$

Lastly stage (iv) gives $r_2(\theta)$. The value, for $R=1$, of the terms of dimension 2 in the coefficient of f_{RR} in (6) is simple, although these terms are so complicated. This is because $f_1=f_{1R}=f_{1\theta}=0$ on $R=1$ and because $r_1(\theta)=0$. (The latter condition implies that no terms in the coefficient of f_{RR} in (16) have dimension 2.) Thus condition (iv) becomes

$$-2r_2 + [-(\gamma+1)M^2\beta^{-2}(f_2 - f_{2R}) - (\gamma-1)M^2(f_2 - f_{2R}) + 2M^2f_{2R}]_{R=1} = 0. \quad (20)$$

When it shall have been shown that $(f_2)_{R=1}=0$, (19) and (20) will give $r_2(\theta)$ as a multiple of $A^2(\theta)$.

The shock conditions will now be applied at a curve $r=\eta(\theta)$, to be determined. By (12), with $f \equiv 0$ ahead of the shock, these say that directly behind it

$$f=0, \quad f_R = \frac{4\beta^2(\eta-1)}{(\gamma+1)M^4} + O[(\eta-1)^2 + \eta'^2]. \quad (21)$$

Since the pressure cannot decrease on passage through a shock, $\eta \geq 1$. But since, if the method is to be applicable, the shock must come in the range $R < 1$, we have $\eta < 1 + r_2 + r_3 + \dots$. Thus it is necessary to find a solution in which $\eta-1$ is of dimension 2 at most, so that $\eta = 1 + \eta_2 + \eta_3 + \dots$, where η_k is of dimension k , and $0 \leq \eta_2 \leq r_2$. Now, in the condition $f=0$ on $r=\eta(\theta)$, the terms f_1, f_2, \dots in f can be expanded by Taylor's theorem (applied at $R=1$), using that $R=1+(\eta_2-r_2)+(\eta_3-r_3)+\dots$, and it is possible to select the terms of each dimension and equate them to zero. Clearly the terms of dimensions 1 and 2 respectively give $(f_1)_{R=0}=0$ and $(f_2)_{R=0}=0$, the first agreeing with what has already been assumed, and the second giving, with (19) and (20), that

$$(f_{2R})_{R=1} = \frac{1}{2}(\gamma+1)M^4\beta^{-2}A^2(\theta), \quad r_2(\theta) = \frac{1}{4}(\gamma+1)^2M^8\beta^{-4}A^2(\theta). \quad (22)$$

To express the second condition of (21), f_R is needed when

$$R = 1 + (\eta_2 - r_2) + (\eta_3 - r_3) + \dots$$

The largest terms therein are of dimension 2, and these come from $f_{1R} \sim A(\theta)\sqrt{1-R}$ and from $(f_{2R})_{R=1}$; thus they are

$$A(\theta)\sqrt{r_2 - \eta_2} + (f_{2R})_{R=1}.$$

Hence, by (21) and (22),

$$A(\theta)\sqrt{\{\frac{1}{4}(\gamma+1)^2M^8\beta^{-4}A^2(\theta)-\eta_2\}+\frac{1}{2}(\gamma+1)M^4\beta^{-2}A^2(\theta)}=4\beta^2\eta_2/(\gamma+1)M^4, \quad (23)$$

whence the goal of this section, the deduction of η_2 (and hence the approximate shock strength) from the given $A(\theta)$, can be achieved.

To solve (23) put $\eta_2=\frac{1}{4}(\gamma+1)^2M^8\beta^{-4}K(\theta)$; it becomes

$$A(\theta)\sqrt{\{A^2(\theta)-K(\theta)\}}=2K(\theta)-A^2(\theta). \quad (24)$$

If (24) is squared, and the resulting quadratic for $K(\theta)$ solved, the roots are 0 and $\frac{3}{4}A(\theta)$. But only one of these can satisfy (24) if, as was specified originally, the positive value of the square root is taken therein. Thus, by inspection, if $A(\theta)>0$, $K(\theta)=\frac{3}{4}A(\theta)$ is the only solution; while if $A(\theta)<0$, $K(\theta)=0$ is the only solution.

In other words when $A(\theta)>0$, so that on linear theory the pressure gradient just behind the Mach cone is positive (and $\rightarrow +\infty$ as $r \rightarrow 1$), there is a shock, given to the first approximation by

$$r=1+\frac{3}{16}(\gamma+1)^2M^8\beta^{-4}A^2(\theta). \quad (25)$$

But when $A(\theta)<0$, so that on linear theory the pressure gradient just behind the Mach cone is *negative* (*i. e.* $\rightarrow -\infty$ as $r \rightarrow 1$), there is no shock (at any rate of strength which is of dimension 2, *i. e.* of order of magnitude the square of the disturbance velocities). It is interesting how the little twist which the mathematics has given in the solution of (24) corresponds so well with what seems physically plausible by the second law of thermodynamics.

To obtain, when $A(\theta)>0$, the relative pressure change at the shock (usually taken as a measure of its strength) observe that from (21) the pressure change is, to a first approximation,

$$\rho_1 U(U-\phi_z)=\rho_1 U^2(rf_r-f)=\rho_1 U^2 4\beta^2(\eta-1)/(\gamma+1)M^4.$$

Since $\rho_1 U^2/p_1=\gamma M^2$, the shock strength reduces to

$$\frac{\Delta p}{p_1} \simeq \frac{4\gamma\beta^2}{(\gamma+1)M^2} \eta_2(\theta) = \frac{3}{4}\gamma(\gamma+1) \frac{M^6}{M^2-1} A^2(\theta), \text{ when } A(\theta)>0; \quad (26)$$

where the reader is reminded that $A(\theta)$ is the value of

$$\lim_{r \rightarrow 1} [(p-p_1)/\rho_1 U^2(1-r)^{\frac{1}{2}}]$$

on linear theory. The form of (26) indicates that the shock strength becomes larger as $M \rightarrow 1$ or as $M \rightarrow \infty$, but also that its value as an approximation probably drops as these limits are approached.

The pressure gradient immediately behind the shock is given by

$$\begin{aligned} \frac{\partial}{\partial r} \left(\frac{p-p_1}{\rho_1 U^2} \right) &\simeq r f_{rr} \simeq -\frac{1}{2} A(\theta) (1-R)^{-\frac{1}{2}} \simeq -\frac{1}{2} A(\theta) (r_2-\eta_2)^{-\frac{1}{2}} \\ &= -\frac{1}{2} A(\theta) / \frac{1}{4} (\gamma+1) M^4 \beta^{-2} A(\theta) = -\frac{2\beta^2}{(\gamma+1)M^4}, \quad (27) \end{aligned}$$

whence (remembering which is the direction r increasing) there is a finite rate of increase of pressure directly behind the shock, which is *not* small, however small be the disturbance in the field. This gradient soon drops, of course, as one goes away from the shock (with $1-R$ increasing), and drops sooner the smaller be the disturbances. The thin region of rapid compression (not of course large enough to bring viscosity into play) "reinforces" the shock (itself a region of almost infinitely rapid compression).

When $A(\theta) < 0$ the shock, if it occurs, has strength of dimension at least 3. To find η_3 would be an arduous process, involving the complications mentioned in footnotes to §3. The author believes that a shock of some strength not zero probably occurs (see the next paragraph). But whatever the strength of the shock, the region of negative non-small pressure gradient behind it is, in a sense, a phenomenon on a larger scale. The gradient immediately behind it is

$$\begin{aligned} \frac{\partial}{\partial r} \left(\frac{p-p_1}{\rho_1 U^2} \right) &\doteq f_{rr} \doteq -\frac{1}{2} A(\theta) (1-R)^{-\frac{1}{2}} = -\frac{1}{2} A(\theta) \{r_2 + (r_3 - \eta_3) + \dots\}^{-\frac{1}{2}} \\ &\doteq -\frac{1}{2} A(\theta) / \{-\frac{1}{2}(\gamma+1)M^4 \beta^{-2} A(\theta)\} = + \frac{\beta^2}{(\gamma+1)M^4} \dots \dots \dots (28) \end{aligned}$$

A gradient of this order is maintained in a region whose width is of dimension 2: it has fallen to about half this value when $r=1-3r_2(\theta)$. It is interesting how in this problem nature simulates a "shock of negative strength" (as she does also in the "Prandtl-Meyer expansions" of plane flow). This would probably be the most noticeable effect on an interferometer photograph, as the pressure change is larger than at the shock. It is this phenomenon which the linear theory crudely represents by a pressure gradient of $-\infty$ at $r=1$.

There are two questions which might be asked, regarding the possibility of a flow with $A(\theta) < 0$ beginning on $r=1$ exactly, which the author can answer. Firstly, it is true that a Taylor expansion at $r=1$ shows, independently of the above theory, that if, at $r=1$, $f=f_r=0$, then $f_{rr}=\beta^2/(\gamma+1)M^4$. However it also shows that f_{rrr} must be logarithmically infinite at $r=1$, a very suspicious circumstance, since this point, where $R=1-r_2-r_3-\dots$, is well within the presumed radius of convergence of the expansions (14) of §3, which should therefore have no singularity there. Probably therefore a flow given by such a Taylor expansion has further singularities for non-small r , which prevent it representing a small disturbance to the main stream. (In any case, judging by the values of the repeated derivatives of f at $r=1$ given by the present theory, it has a very small radius of convergence). If this conclusion is correct, there must be a shock.

Secondly, is there (as in a Prandtl-Meyer expansion) a second characteristic of discontinuity, inside $r=1$, whose choice might introduce enough generality to prevent the necessity for a shock? No; equation (6) is elliptic in $r < 1$, because the coefficient of f_{rr} therein is positive. For its

derivative with respect to r contains two terms only which are of dimension zero, namely $-2r + (\gamma + 1)M^2(\beta^{-2}r^3 + r)f_{rr}$, and the latter term has a maximum of approximately unity, attained when $r=1$. Hence the coefficient of f_{rr} increases from zero, as r decreases from unity, at least when say $r > \frac{3}{4}$; but is clearly positive for $r < \frac{3}{4}$, assuming small disturbances. Crudely, the rapid expansion rate (28) tries to keep the equation hyperbolic for a time, but fails by half.

Examples on the above theory are deferred to § 6.

§ 5. DETERMINATION OF THE SHOCK STRENGTH IN THE GENERAL CASE.

When part of the solid boundary lies outside the Mach cone from the apex of the conical field, then the disturbance velocities are not zero in the region $r > 1$, and a linearized solution must proceed first by determining the particular solutions of the wave equation in the variables $\sec^{-1} r$ and θ (for the velocity components ϕ_x , ϕ_y , and ϕ_z) which correspond to the given boundary conditions for $r > 1$; and secondly by using the resulting values of the velocities on $r=1$, and further boundary conditions, to determine the components inside the circle, where they satisfy Laplace's equation in $\text{sech}^{-1} r$ and θ .

As $r \rightarrow 1$, the linearized theory therefore gives for f_r and f a behaviour of the form

$$\left. \begin{aligned} f_r &= C(\theta) + B(\theta)(r-1)^{\frac{1}{2}} + O(r-1), & (r > 1), \\ f_r &= C(\theta) + A(\theta)(1-r)^{\frac{1}{2}} + O(1-r), & (r < 1), \\ f &= D(\theta) + C(\theta)(r-1) + O(|r-1|^{\frac{3}{2}}), & (r \gtrless 1). \end{aligned} \right\} \quad \dots \quad (29)$$

Carrying out stage (i) of the process of § 3 gives an $f_1(R, \theta)$ which is exactly the linearized $f(r, \theta)$ with R for r , and so satisfies (29) with f_1 for f and R for r .

Carrying out stage (ii), that is choosing $r_1(\theta)$ in (14) to make the terms of dimension 1 in the coefficient of f_{RR} in (6) vanish when $R=1$, gives that

$$-2r_1 + (\gamma + 1)M^4\beta^{-2}C(\theta) - ((\gamma + 1)\beta^{-2} + \gamma - 1)M^2D(\theta) = 0. \quad \dots \quad (30)$$

Carrying out stage (iii), for $R=1$ only, gives

$$\begin{aligned} \lim_{R \rightarrow 1} \{ (\gamma + 1)M^4\beta^{-2}f_{1R}f_{1RR} \} + (f_{2R} + f_{2\theta\theta})_{R=1} - \{ r_1(\theta) + (\gamma - 1)M^2(D(\theta) \\ - C(\theta)) \} C(\theta) - \{ 2r_1(\theta) + (\gamma - 1)M^2(D(\theta) - C(\theta)) \} D''(\theta) \\ - 2M^2(D'(\theta) - C'(\theta))D'(\theta) = 0. \quad \dots \quad (31) \end{aligned}$$

The limit in (31) is $\frac{1}{2}(\gamma + 1)M^4\beta^{-2}B^2(\theta)$ or $-\frac{1}{2}(\gamma + 1)M^4\beta^{-2}A^2(\theta)$ according as $R \rightarrow 1$ from above or from below. Thus $f_{2R} + f_{2\theta\theta}$ has a different limit according as $R \rightarrow 1$ from above or from below, and the difference of the former limit from the latter; which may be written $\Delta(f_{2R} + f_{2\theta\theta})$, is

$$\Delta(f_{2R} + f_{2\theta\theta}) = \frac{1}{2}(\gamma + 1)M^4\beta^{-2}(A^2(\theta) + B^2(\theta)). \quad \dots \quad (32)$$

Lastly, carrying out stage (iv) gives a rather lengthy expression for r_2 (in which terms from (16) must be taken into account besides terms from the coefficient of f_{rr} in (6)). However it simplifies greatly if the terms in f_1 are not written down explicitly, when it becomes

$$-2r_2 + (\gamma + 1)M^4\beta^{-2}(f_{2R})_{R=1} - ((\gamma + 1)\beta^{-2} + \gamma - 1)M^2(f_2)_{R=1} \\ = \text{a quadratic form in } C(\theta), C'(\theta), D(\theta) \text{ and } D'(\theta). \quad (33)$$

(The term in $C'^2(\theta)$ on the right arises from the term in $r_1'^2$ in (16), and indeed is $-\frac{1}{4}(\gamma + 1)^2M^8\beta^{-4}C'^2(\theta)$.) Equations (32) and (33) indicate that the function $r_2(\theta)$ will be different in the solutions for $R > 1$ and $R < 1$ respectively. These correspond to $r > 1 + r_1(\theta) + r_2(\theta) + \dots$ and $r < 1 + r_1(\theta) + r_2(\theta) + \dots$, with the same value of $r_1(\theta)$ in each but different values of $r_2(\theta)$. The shock $r = \eta(\theta)$ must come in the region of overlap, and therefore $\eta(\theta) = 1 + r_1(\theta) + \eta_2(\theta) + \eta_3(\theta) + \dots$, where $\eta_2(\theta)$ lies between the two $r_2(\theta)$'s. The value of R at the shock is $1 + (\eta_2 - r_2) + (\eta_3 - r_3) + \dots$

By the last equation of (29) the change in f_1 at the shock is

$$\Delta f_1 = O\{C(\theta)\Delta R + \Delta|R - 1|^{\frac{3}{2}}\}, \quad (34)$$

which is of dimension 3 at most. Since $\Delta f = \Delta(f_1 + f_2 + \dots) = 0$ at the shock, Δf_2 is also of dimension 3 at most. Hence $(f_2)_{R=1}$ has the same value in the two regions $R \geq 1$ and $R \leq 1$, and by (32) and (33)

$$\Delta f_{2R} = \frac{1}{2}(\gamma + 1)M^4\beta^{-2}(A^2(\theta) + B^2(\theta)), \quad \Delta r_2 = \frac{1}{4}(\gamma + 1)^2M^8\beta^{-4}(A^2(\theta) + B^2(\theta)). \quad (35)$$

Since Δr_2 is positive, the two regions do overlap. Without knowing η_2 or r_2 precisely one may now put for $\eta_2 - r_2$, the term of dimension 2 in the value of R at the shock,

$$\eta_2 - r_2 = \frac{1}{4}(\gamma + 1)^2M^8\beta^{-4}(F(\theta) + B^2(\theta)), \quad (R > 1), \\ = \frac{1}{4}(\gamma + 1)^2M^8\beta^{-4}(F(\theta) - A^2(\theta)), \quad (R < 1), \quad (36)$$

for some $F(\theta)$.

The condition on Δf_r at the shock will now be expressed. To simplify matters, it will be supposed that, in applying the condition for $\theta = \theta_0$, the axes and the value of U are specially chosen so that $C(\theta_0) = D(\theta_0) = D'(\theta_0) = 0$, *i. e.* so that there is no "disturbance velocity" on linear theory at $r = 1$, $\theta = \theta_0$. This requires simply that the "undisturbed velocity" should be taken equal in magnitude and direction to the velocity at this point. The Mach cone from the apex will shift in consequence, but on a linear theory this will not affect the velocity at $r = 1$, $\theta = \theta_0$. At any rate it seemed plausible that the change would only alter the velocity field by a constant and that therefore the shock strength deduced below entirely in terms of $A(\theta)$ and $B(\theta)$ would be correct, whether these were calculated before or after the transformation. However, in order to be certain that this is so, the author has gone through the lengthy process of checking that the same conclusion is reached without restriction on the values of $C(\theta_0)$, $D(\theta_0)$ and $D'(\theta_0)$. It is; but in this paper only the work with these equated to zero will be exhibited.

Consequences are that $r_1(\theta_0)=0$ and that the right-hand side of (33) vanishes except for the term in $C'^2(\theta_0)$ noted beneath it. However $r'_1(\theta_0)$ is not zero in general but is $\frac{1}{2}(\gamma+1)M^4\beta^{-2}C'(\theta_0)$ by (30). Hence in the shock condition (12) it is permissible to neglect $(\eta-1)^2$ and the squares of the disturbance velocities, but not η'^2 , which is of dimension 2. This gives an additional term $\beta^2\eta'^2$ in the denominator $\beta^2+\eta'^2$ in (11), and hence an additional term $-M^{-4}\beta^2\eta'^2$ within the square brackets of (12). Hence condition (12) (ii) becomes, using (29), (36), and (35),

$$\begin{aligned} A(\theta_0)\frac{1}{2}(\gamma+1)M^4\beta^{-2}(A^2(\theta_0)-F(\theta_0))^{\frac{1}{2}}-B(\theta_0)\frac{1}{2}(\gamma+1)M^4\beta^{-2}(B^2(\theta_0)+F(\theta_0))^{\frac{1}{2}} \\ +\frac{1}{2}(\gamma+1)M^4\beta^{-2}(A^2(\theta_0)+B^2(\theta_0))=\frac{2}{\gamma+1}[\{(\gamma-1+2M^{-2})f_2-(\gamma+1)f_{2R}\}_{R=1} \\ -(\gamma+1)B(\theta_0)\frac{1}{2}(\gamma+1)M^4\beta^{-2}(B^2(\theta_0)+F(\theta_0))^{\frac{1}{2}} \\ +M^{-4}\beta^2(2\eta_2-\frac{1}{4}(\gamma+1)^2M^8\beta^{-4}C^2(\theta_0))] \\ =-B(\theta_0)(\gamma+1)M^4\beta^{-2}(B^2(\theta_0)+F(\theta_0))^{\frac{1}{2}} \\ +\frac{2}{\gamma+1}M^{-4}\beta^2[2\eta_2-\frac{1}{4}(\gamma+1)^2M^8\beta^{-4}C^2(\theta_0) \\ +((\gamma+1)\beta^{-2}+\gamma-1)M^2(f_2)_{R=1}-(\gamma+1)M^4\beta^{-2}(f_{2R})_{R=1}]. \quad \dots \quad (37) \end{aligned}$$

By (33) and (36) the term in square brackets on the right is

$$2\eta_2-2r_2=\frac{1}{2}(\gamma+1)^2M^8\beta^{-4}(F(\theta_0)+B^2(\theta_0)).$$

Hence (37) becomes the following equation for F , which as mentioned above has been also checked by the author without the assumption that C , D and D' vanish at the point under consideration, and which is therefore written with the zero suffix dropped from θ :

$$A(\theta)(A^2(\theta)-F(\theta))^{\frac{1}{2}}+B(\theta)(B^2(\theta)+F(\theta))^{\frac{1}{2}}+A^2(\theta)-B^2(\theta)=2F(\theta). \quad (38)$$

(This reduces to the equation (24) for $K(\theta)$, in §4, when $B(\theta)=0$).

The quantity F , defined in (36) and evaluable from (38), does not specify the position of the shock to dimension 2 (which to dimension 1 is $r=1+r_1$ with r_1 given from (30)). This cannot be deduced merely from the behaviour of the linearized solution. However, the *strength* is deducible from the value of R at the shock, which is known to dimension 2 if F is known. In fact the shock strength to dimension 2 is

$$\begin{aligned} \frac{\Delta p}{p_1}=\frac{\rho_1 U^2 \Delta f_r}{p_1}=\gamma M^2 \Delta f_r=\frac{1}{2}\gamma(\gamma+1)M^6\beta^{-2} \\ \times [A(\theta)(A^2(\theta)-F(\theta))^{\frac{1}{2}}-B(\theta)(B^2(\theta)+F(\theta))^{\frac{1}{2}}+A^2(\theta)+B^2(\theta)]. \quad (39) \end{aligned}$$

The right-hand side is essentially positive for any set of numbers $A(\theta)$, $B(\theta)$, $F(\theta)$. By (38) it can also be written

$$\gamma(\gamma+1)M^6\beta^{-2}[F(\theta)+B^2(\theta)-B(\theta)(B^2(\theta)+F(\theta))^{\frac{1}{2}}].$$

Of course in §4 position as well as strength was calculated to dimension 2, but there each followed from the other, since the flow ahead of the shock was *exactly* known.

If the surds in (38) are removed by judicious squaring, a quartic in F results, of which one root is $F=0$. Only one of its four roots will actually solve (38), if the square roots therein are taken both positive (similarly only one would solve it if any other distribution of sign between the two were specified): clearly the other roots will solve (38) with $-A$ for A , or $-B$ for B , or both. The root $F=0$ solves (38) when $A(\theta)\leq 0$ and $B(\theta)\geq 0$.

Since however $F(\theta)$ is only interesting as an auxiliary function, whence the strength is determined, only the corresponding quartic for the strength will be displayed. Putting

$$\frac{4p}{p_1} = \gamma(\gamma+1)M^6\beta^{-2}S(\theta) + \text{terms of dimension } >2, \quad . \quad . \quad (40)$$

we have

$$S = F + B^2 - B\sqrt{(B^2 + F)} = A\sqrt{(A^2 - F)} + A^2 - F.$$

Hence, $F = S - \frac{1}{2}B^2 + \frac{1}{2}B\sqrt{(B^2 + 4S)}$, with positive square root (necessary since it has been proved that $S \geq 0$), and

$$S = A\sqrt{\{A^2 - S + \frac{1}{2}B^2 - \frac{1}{2}B\sqrt{(B^2 + 4S)}\}} + A^2 - S + \frac{1}{2}B^2 - \frac{1}{2}B\sqrt{(B^2 + 4S)}. \quad (41)$$

The quartic derived from (41) is

$$16S^4 - 24(A^2 + B^2)S^3 + 9(A^2 + B^2)^2S^2 - 4(A^2 + B^2)A^2B^2S = 0. \quad (42)$$

One root is $S=0$, which solves (41) when $A \leq 0$ and $B \geq 0$. The others are positive (when neither A nor B vanishes) and add up to $\frac{3}{2}(A^2 + B^2)$. (When $B=0$ as in §4, the roots are $\frac{3}{4}A^2$, $\frac{3}{4}A^2$, 0, 0.) The cubic, for the three roots other than zero, is most easily solved by putting $A = H \cos \lambda$, $B = H \sin \lambda$, $S = H^2 \sin^2 \psi$, where $H > 0$; when it becomes simply $\sin^2 3\psi = \sin^2 2\lambda$. One may take $\psi = \frac{2}{3}\lambda$ for each of the three roots, the values of λ differing by multiples of 2π which leave A and B unaltered. The appropriate root which actually solves (41) depends on the quadrant in which the angle finds itself. One finds the following rule for determining it.

$$\left. \begin{aligned} \text{1st Quadrant } (A \geq 0, B \geq 0) : S &= H^2 \sin^2 \left(\frac{2}{3}\lambda\right) \text{ where } -2\pi \leq \lambda \leq -\frac{3}{2}\pi; \\ \text{2nd Quadrant } (A \leq 0, B \geq 0) : S &= 0; \\ \text{3rd Quadrant } (A \leq 0, B \leq 0) : S &= H^2 \sin^2 \left(\frac{2}{3}\lambda\right) \text{ where } -3\pi \leq \lambda \leq -\frac{5}{2}\pi; \\ \text{4th Quadrant } (A \geq 0, B \leq 0) : S &= H^2 \sin^2 \left(\frac{2}{3}\lambda\right) \text{ where } -\frac{5}{2}\pi \leq \lambda \leq -2\pi. \end{aligned} \right\} \quad (43)$$

(Here $A = H \cos \lambda$, $B = H \sin \lambda$). It will be observed that in the first and third quadrants S cannot exceed $\frac{3}{4}(A^2 + B^2)$, while in the fourth it must equal or exceed it.

Physical meaning can be given to these results as follows. Regard the shock as formed from the interaction of two effects: the shock of §4, needed to get the required compression inside the Mach cone, with $B=0$ and $S = \frac{3}{4}A^2$ (or $S=0$ when $A \leq 0$); and a shock, with $A=0$, due to the

“ piling up of characteristics ” outside the Mach cone : this has $S = \frac{3}{4}B^2$ (or $S=0$ when $B \geq 0$). Interaction of the two shocks of positive strengths gives a strength exceeding their sum. Interaction of either with one of the zero-strength shocks reduces its strength*, but not to zero. When the two expansions interact, the shock strength has dimension at most 3 : and it is an easy matter to show that the pressure gradient $\partial(p/\rho_1 U^2)/\partial r$ is $\beta^2/(\gamma+1)M^4$ at the shock (as in (28)), and falls to half this value at two points, one approximately $\frac{3}{4}(\gamma+1)^2 M^8 \beta^{-4} B^2(\theta)$ ahead of it, and the other $\frac{3}{4}(\gamma+1)^2 M^8 \beta^{-4} A^2(\theta)$ behind it.

Summing up, the position of the shock is given to dimension 1 by

$$r = 1 + r_1(\theta) = 1 + \frac{1}{2}(\gamma+1)M^4\beta^{-2}C(\theta) - \frac{1}{2}((\gamma+1)\beta^{-2} + \gamma - 1)M^2D(\theta) \\ = 1 + [(M^2/\beta U)(\phi_x \cos \theta + \phi_y \sin \theta) - (1 + \frac{1}{2}(\gamma-1)M^2)(\phi_z - U)/\beta^2 U]_{r=1}, \quad (44)$$

the curve on which, to this dimension, the character of equation (6) changes from hyperbolic to elliptic ; its strength, which is of dimension 2, is given by equation (40), in which $S(\theta)$ is to be determined from (43) ; here $A(\theta)$, $B(\theta)$ are the coefficients of $\sqrt{1-r}$, $\sqrt{r-1}$ respectively in the behaviour of $(p-p_1)/\rho_1 U^2$ as $r \rightarrow 1$ from below and above.

§ 6. EXAMPLES.

As a first example of the theory of § 4, consider the flow past any cone (in the general sense), which is slender in all directions perpendicular to the main stream, but whose cross-section is a curve without any points of abnormally large curvature. Ward (1949) has investigated on a linear theory, the flow past *any* pointed body which is slender in this sense. He uses the operational calculus, with p as Heaviside's equivalent of $\partial/\partial z$. Then (Ward, 1949, equation (7))

$$U^{-1}\phi - z = A_0 K_0(\beta p \rho) + \sum_{n=1}^{\infty} A_n K_n(\beta p \rho) \cos(n\theta + \epsilon_n), \quad \dots \quad (45)$$

where the A_n and ϵ_n depend on p and are to be determined from the boundary condition at the surface. If the cross-section of the body normal to the stream at a distance z from the pointed nose is $S(z)$, then (Ward 1949, equations (11) and (26)) A_0 must be the operational form of $-S'(z)/2\pi$. In the case of a slender cone of solid angle Ω , $S(z) = \Omega z^2$. Hence

$$A_0 = -\frac{\Omega}{\pi p}. \quad \dots \dots \dots (46)$$

But in fact ϕ_z must depend only on $r = \beta p/z$ and on θ ; hence its operational form $p\phi$ depends only on $\beta p \rho$ and θ , and each A_n (like A_0) is a mere constant a_n divided by p , while ϵ_n is independent of p . Using these facts, the asymptotic form of (45) as $p\rho \rightarrow \infty$ (Ward (1949), equation (8)) is

$$zf(r, \theta) \sim \left[-\frac{\Omega}{\pi} + \sum_{n=1}^{\infty} a_n \cos(n\theta + \epsilon_n) \right] \frac{1}{p} \left(\frac{\pi}{2\beta p \rho} \right)^{\frac{1}{2}} e^{-\beta p \rho}, \quad \dots \quad (47)$$

* e. g. in the first quadrant $S = H^2 \sin^2(\frac{2}{3}\lambda) \leq \frac{3}{4}H^2 \cos^2 \lambda = \frac{3}{4}A^2$.

which on interpretation becomes

$$\begin{aligned} f(r, \theta) &\sim \left[-\frac{\Omega}{\pi} + \sum_{n=1}^{\infty} a_n \cos(n\theta + \epsilon_n) \right] \frac{(z - \beta\rho)^{\frac{1}{2}}}{\left(\frac{3}{2}\right)! z} \left(\frac{\pi}{2\beta\rho}\right)^{\frac{1}{2}} \\ &= \left[-\frac{\Omega}{\pi} + \sum_{n=1}^{\infty} a_n \cos(n\theta + \epsilon_n) \right] \frac{1}{2} 2^{\frac{1}{2}} (1-r)^{\frac{1}{2}} r^{-\frac{1}{2}}. \quad \dots (48) \end{aligned}$$

Hence, in the theory of this paper,

$$A(\theta) = \lim_{r \rightarrow 1} \frac{f_r}{(1-r)^{\frac{1}{2}}} = \sqrt{2} \left[\frac{\Omega}{\pi} - \sum_{n=1}^{\infty} a_n \cos(n\theta + \epsilon_n) \right]. \quad \dots (49)$$

Now the series in (49) is small compared with the constant term Ω/π . For interpretation of equations (9) and (22) of Ward (1949) gives that his function

$$\phi_0 = \frac{\Omega}{\pi} z \log \frac{2z}{e\beta\rho} + \frac{1}{2} z \sum_{n=1}^{\infty} \left(\frac{2z}{\beta\rho} \right)^n \frac{a_n \cos(n\theta + \epsilon_n)}{n(n+1)}, \quad \dots (50)$$

must satisfy

$$\frac{\partial \phi_0}{\partial \nu} = \frac{d\nu}{dz} \quad \dots (51)$$

at the boundary of the cross-section which is at distance z from the nose, if $\partial/\partial \nu$ denotes differentiation along the outward normal to this boundary, and $d\nu/dz$ is the rate of normal expansion of the boundary with respect to z . For a cone the said rate is independent of z , and is $O(\delta)$, if δ represents the maximum angle to the main stream of any generator. Since the condition (51) on the function (50) is to be applied on a surface whereon $\rho/z = O(\delta)$, the conclusion $a_n = O(\delta^{n+2})$ easily follows: hence every a_n with $n \geq 1$ is of smaller order than $\Omega = O(\delta^2)$.

Thus, by (25) and (26), for supersonic flow past a cone, all of whose generators make a small angle with the main stream, the shock is approximately a right circular cone

$$r = 1 + \frac{3}{8}(\gamma + 1)^2 \frac{M^8}{(M^2 - 1)^2} \frac{\Omega^2}{\pi^2}, \quad \dots (52)$$

with axis in the stream-direction, and its strength is approximately uniform and equal to

$$\frac{3}{2}\gamma(\gamma + 1) \frac{M^6}{M^2 - 1} \frac{\Omega^2}{\pi^2}, \quad \dots (53)$$

where M is the Mach number of the main stream and Ω the solid angle subtended by the cone. In particular the rather remarkable fact emerges that a change of incidence, not large enough to increase materially the maximum angle made by a generator with the main stream, will not substantially alter the shock strength or position.

As a second example of the theory of §4, consider the supersonic flow past a flat triangular wing, at incidence α , which lies entirely inside the Mach cone from one of its vertices (the apex of the conical field) and outside those from the other two. Suppose that it lies in the (x, z) plane when $\alpha = 0$ and that increase of α means rotation in the positive sense about the x -axis. If its two leading edges make angles τ_0, τ_1 with the x -axis, then

$\tan^{-1} \beta < \tau_0 < \tau_1 < \pi - \tan^{-1} \beta$ has been assumed already; suppose also that $\tau_0 < \frac{1}{2}\pi < \tau_1$. The numbers $t_0 = \beta \cot \tau_0$ and $t_1 = \beta \cot \tau_1$ then satisfy $-1 < t_1 < 0 < t_0 < 1$. The linearized theory of the flow is given by Goldstein and Ward (1949) in their §11. The disturbance velocity $\phi_z - U$ is a harmonic function of θ and of $\text{sech}^{-1} r$, which latter they write as $-s$; hence it is the real part of some regular function of $s + i\theta$, say $G_3(s + i\theta)$. It is clearly antisymmetrical between the two semicircles $0 < \theta < \pi$ (above the wing) and $\pi < \theta < 2\pi$ (below it). By the conformal transformation $t = \text{sech}(s + i\theta)$ the upper semi-circle becomes the upper half t -plane; G_3 must be a regular function of t , and in fact (Goldstein and Ward (1949), equation (127))

$$G_3'(t) = - \frac{iLU\alpha t}{\beta(t-t_0)^{\frac{1}{2}}(t-t_1)^{\frac{1}{2}}}, \quad \dots \dots \dots (54),$$

where

$$\left. \begin{aligned} L &= \frac{(t_0 - t_1)^2}{2(1+t_0)^{\frac{1}{2}}(1-t_1)^{\frac{1}{2}}(2E(k) - (1-k^2)K(k))}, \\ k &= [(1-t_0)(1+t_1)/(1+t_0)(1-t_1)]^{\frac{1}{2}}, \end{aligned} \right\} \quad \dots \dots (55).$$

and $E(k)$, $K(k)$ signify complete elliptic integrals in the usual notation..

Now, in the theory of this paper, $A(\theta)$ is given in terms of the linearized potential ϕ as

$$\begin{aligned} A(\theta) &= \lim_{r \rightarrow 1} \left[\frac{1 - U^{-1}\phi_z}{(1-r)^{\frac{1}{2}}} \right] = \lim_{r \rightarrow 1} \left[\frac{-U^{-1}RG_3}{2^{-\frac{1}{2}} \text{sech}^{-1} r} \right] = \frac{\sqrt{2}}{U} \left(\frac{dRG_3}{ds} \right)_{s=0} \\ &= \frac{\sqrt{2}}{U} \left(\frac{dG_3}{ds} \right)_{s=0} \quad \dots \dots \dots (56), \end{aligned}$$

since RG_3 itself vanishes on $s=0$. Hence

$$\begin{aligned} A(\theta) &= \frac{\sqrt{2}}{U} \left(\frac{dt}{ds} \right)_{s=0} G_3'(\sec \theta) = \frac{\sqrt{2}}{U} \left(- \frac{i \sin \theta}{\cos^2 \theta} \right) \left(- \frac{iLU\alpha \sec \theta}{\beta(\sec \theta - t_0)^{\frac{1}{2}}(\sec \theta - t_1)^{\frac{1}{2}}} \right) \\ &= - \frac{L\alpha\sqrt{2} \sin \theta}{\beta(1-t_0 \cos \theta)^{\frac{1}{2}}(1-t_1 \cos \theta)^{\frac{1}{2}}}, \quad \dots \dots \dots (57), \end{aligned}$$

for $0 < \theta < \pi$; but by antisymmetry the same formula holds for $\pi < \theta < 2\pi$. Hence, by (26), there is in $\pi < \theta < 2\pi$ a shock strength varying with θ approximately like

$$\frac{3}{2}L^2\gamma(\gamma+1) \frac{M^6}{(M^2-1)^2} \alpha^2 \frac{\sin^2 \theta}{(1-t_0 \cos \theta)^3(1-t_1 \cos \theta)^3}; \quad \dots \dots (58).$$

but in $0 < \theta < \pi$ the strength is of order smaller than the square of the angle of incidence, and the principal phenomenon is a rapid expansion. The region where the shock strength is of order α^2 is the half of the Mach cone “ below ” the plane of the wing if the nose is turned “ up ”. Expression (58) has double zeros at $\theta=0, \pi$ where the regions meet (this would happen in any example of the theory of §4), giving a smooth transition. To get some idea of its numerical magnitude, let $M=\sqrt{2}$, $\gamma=1.4$, $\tau_0=\frac{3}{8}\pi$, $\tau_1=\frac{5}{8}\pi$. Then the leading edges are both swept back at an angle 67.5° to the main stream, $t_0=-t_1=\tan \frac{1}{8}\pi=0.4142$, and (58) has a maximum value (at

$\theta = \frac{3}{2}\pi$) of $0.866\alpha^2$, which means that at 10° incidence the maximum shock strength is 0.026. As t_0 approached 1 and t_1 approached -1 this would increase. If $t_0 = -t_1$ then, when $t_0 > 1/\sqrt{3}$, the maximum of (58) is no longer at $\theta = \frac{3}{2}\pi$ but at two points satisfying $2t_0^2 \cos^2 \theta = 3t_0^2 - 1$; the shock strength then rises to a maximum on two symmetrically placed generators, and the central generator of the shock is a local minimum of strength.

Passing on to the theory of §5, it must be regretted that no example of the full theory can be given, because, in all solutions of the linearized conical field equations which have come to the author's notice, the field outside the Mach cone, though disturbed, is divided into a number of regions in each of which the flow is uniform. Hence the function $B(\theta)$ of §5 never takes any value other than zero (though there are isolated points on $r=1$ where the linearized velocity field is discontinuous and the meaning of $B(\theta)$ is doubtful). However, even if no particular case of the full theory be ever worked out, the author does not regret having elaborated it, since the physical interpretation of the results, given at the end of §5, is so suggestive and interesting.

As an example of the theory of §5, although with $B(\theta)=0$, the supersonic flow past a flat rectangular wing at incidence is considered. There is a conical field with each end of the leading edge as apex; the interaction of the two fields is not here considered. Choosing axes (for one of the conical fields) so that in the (r, θ) plane the wing takes up the whole of the line $\theta=0$, the problem is the same as that just considered but with $\tau_0=0$, $\tau_1=\frac{1}{2}\pi$, a case included among those considered by Goldstein and Ward (1949) in their §10. There is a region of uniformly disturbed flow, outside the Mach cone and above the wing, specified by the inequalities $\rho > 1$ and $0 < \rho \sin \theta < 1$, where $\phi_x=0$, $\phi_y=-U\alpha$, $\phi_z=U(1+\alpha/\beta)$; and another specified by $\rho > 1$ and $-1 < \rho \sin \theta < 0$ (below the wing), where $\phi_x=0$, $\phi_y=-U\alpha$, $\phi_z=U(1-\alpha/\beta)$. Inside $\rho=1$, in the notation used above,

$$G_3(t) = \frac{2U\alpha}{\beta\pi} \sin^{-1} \sqrt{t}, \quad \dots \dots \dots (59)$$

which gives

$$\begin{aligned} A(\theta) &= \frac{\sqrt{2}}{U} \left(\frac{dt}{ds} \right) G'_3(\sec \theta) \\ &= \frac{\sqrt{2}}{U} \left(-\frac{i \sin \theta}{\cos^2 \theta} \right) \frac{2U\alpha}{\beta\pi} \frac{i}{2[\sec \theta (\sec \theta - 1)]^{\frac{1}{2}}} = \frac{\alpha \tan \theta}{\beta\pi \sin^{\frac{1}{2}} \theta}, \quad \dots \dots (60) \end{aligned}$$

in $0 < \theta < \pi$, and hence also by antisymmetry in $\pi < \theta < 2\pi$.

Hence by the theory of §5 the strength of the shock is

$$\frac{3\gamma(\gamma+1)M^6}{4\pi^2(M^2-1)^2} \alpha^2 \frac{\tan^2 \theta}{\sin^2 \frac{1}{2}\theta}, \quad \dots \dots \dots (61)$$

in $0 < \theta < \frac{1}{2}\pi$ and $\pi < \theta < \frac{3}{2}\pi$, and is of smaller order than α^2 elsewhere. As $\theta \rightarrow 0$ this expression has a finite limit, but this does not mean a discontinuity in shock strength, as the solid wing separates the shocks on either side of $\theta=0$. At $\theta=\pi$, the coefficient of α^2 becomes zero, giving

continuous shock strength to order α^2 at this point. However, at $\theta = \frac{1}{2}\pi$ and $\frac{3}{2}\pi$, (61) becomes infinite, and the significance of this must be carefully considered.

The position of the shock cone, by (44), is

$$r = \begin{cases} 1 - M^2 \alpha \beta^{-1} \sin \theta - (1 + \frac{1}{2}(\gamma - 1)M^2) \alpha \beta^{-3} & (0 < \theta < \frac{1}{2}\pi), \\ 1 & (\frac{1}{2}\pi < \theta < \frac{3}{2}\pi), \\ 1 - M^2 \alpha \beta^{-1} \sin \theta + (1 + \frac{1}{2}(\gamma - 1)M^2) \alpha \beta^{-3} & (\frac{3}{2}\pi < \theta < 2\pi). \end{cases} \quad (62)$$

This must be compared with the position of the plane shock, beneath the wing, which is required to deflect the flow there through an angle α , namely

$$r \sin \theta = -1 - \frac{1}{2}(\gamma + 1)M^4 \beta^{-3} \alpha, \quad \dots \dots \dots (63)$$

which can be deduced from two-dimensional theory or from the formulæ of §2, and with the position of the Prandtl-Meyer expansion required to deflect the flow above the wing through an angle α ; the latter lies between the two lines $r \sin \theta = 1$ and

$$r \sin \theta = 1 - \frac{1}{2}(\gamma + 1)M^4 \beta^{-3} \alpha. \quad \dots \dots \dots (64)$$

Expressions (62), (63), (64) are correct if α^2 be neglected.

Near the discontinuities on $r=1$ at $\theta = \frac{1}{2}\pi$, $\frac{3}{2}\pi$ it must be expected that the theory of §5 does not strictly apply, since it did not allow for such discontinuities; one must consider what in fact happens. The shock (62) intersects the shock (63) where

$$\sin \theta = -1 + \frac{1}{2}\alpha[(3 - \gamma)\beta + (\gamma + 1)\beta^{-3}] + O(\alpha^2), \quad \dots \dots (65)$$

and intersects the lower arm (64) of the Prandtl-Meyer expansion where

$$\sin \theta = 1 - \alpha \beta^{-1}[1 + \frac{1}{2}(\gamma - 1)M^2] + O(\alpha^2), \quad \dots \dots \dots (66)$$

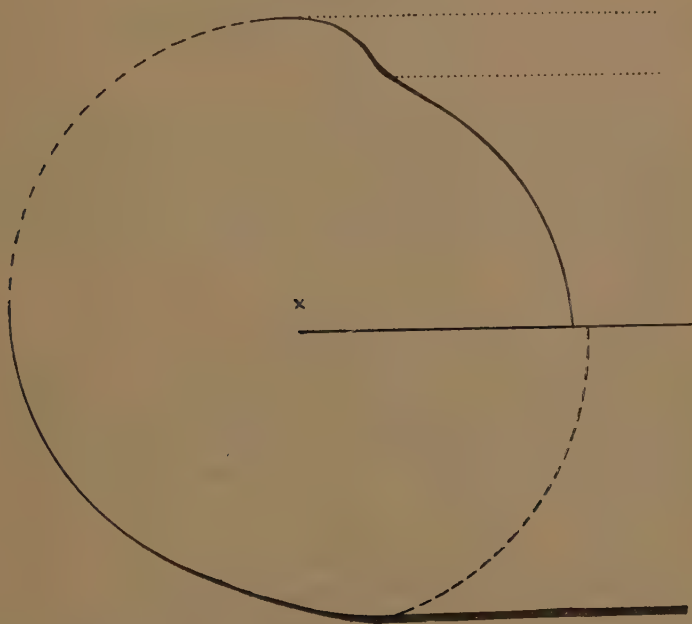
so that in neither case does θ reach the point at which (61) becomes infinite before the shock is interfered with by a phenomenon not allowed for in the theory of §5. It therefore seems clear to the author that the true picture must be as in fig. 1 (where positions of curves are shown correct to order α only). The two arms of the Prandtl-Meyer expansion are shown as dotted lines; the shock (63), whose strength is of order α , is shown as a thick line; the wing, and the shocks whose strengths are of order α^2 , are shown as thin lines; and the shocks of still smaller strength as broken lines. A cross indicates the centre of the undisturbed Mach cone. The shock in $0 < \theta < \frac{1}{2}\pi$ increases in strength with θ , by (61), and (66) indicates that its strength has even reached order α by the time it encounters the Prandtl-Meyer expansion; however in interacting with this it is reduced to negligible proportions (while the expansion itself is reduced in scale from order α to order α^2) before θ exceeds $\frac{1}{2}\pi$ by much*. Again, the expansion in $\frac{3}{2}\pi < \theta < 2\pi$ increases in strength with θ until it meets the shock (63), after which both are reduced in strength and the expansion soon disappears altogether. The shock strength is probably still of order α at the point $\theta = \frac{3}{2}\pi$, since the coefficient of α^2 in the expression

* In this region it is drawn roughly along a characteristic of the Prandtl-Meyer flow.

(61) for it is infinite there. It reduces to order α^2 however when θ has passed $\frac{3}{2}\pi$ by sufficiently much (about $\sqrt{\alpha}$) and thereafter is given adequately by (61), which drops to zero at $\theta=\pi$.

The above is all that can be said without a special investigation near $r=1, \theta=\frac{1}{2}\pi$ and $r=1, \theta=\frac{3}{2}\pi$, which would render the theory of §5 uniformly valid even near these points. The above is sufficient to satisfy the present

Fig. 1.



Shock pattern at tip of rectangular wing in steady supersonic flow.

writer's curiosity ; but he would suggest the further investigation, which he believes would be possible, using an extension of the method of §3, to anyone who may be interested. Similar phenomena to those just discussed and illustrated in fig. 1, will occur in the real solution whenever the linearised velocity field contains lines of discontinuity. When this is not so, the theory of §§4 and 5 provides a complete analysis.

REFERENCES.

- BUSEMANN, A., 1943, "Infinitesimale kegelige Überschallströmung", *Jahrb. d. Luftfahrtf.*, **7B**, 105-121.
 GOLDSTEIN, S., and WARD, G. N., 1949, "The linearized theory of conical fields in supersonic flow, with applications to plane aerofoils", *Aeronautica Quarterly* (to appear).
 LIGHTHILL, M. J., 1948, "The position of the shock-wave in certain aerodynamic problems", *Q.J.M.A.M.*, **1**, 309-318; 1949, "A technique for rendering approximate solutions to physical problems uniformly valid", *Phil. Mag.* [7], **40**, 1179.
 WARD, G. N., 1949, "Supersonic flow past slender pointed bodies", *Q.J.M.A.M.*, **2**, 75-97.

CXI. *A Note on the Identity of Thermal Noise and Shot Noise.*

By B. MELTZER *.

[Received August 31, 1949.]

ABSTRACT.

It is shown that the noise in a given frequency range in an ohmic conductor, originally calculated by Nyquist, can be rigorously treated as pure shot noise. Incidentally, a useful new expression kT/Re is obtained, which gives the average direct current in either direction in a circuit of resistance R in temperature equilibrium, if current is carried by elementary charges of magnitude e .

THE identity of thermal and shot noise in electrical conductors still appears to be doubted by some workers in this field (Moullin 1938, North and others 1940, Bull 1947). The remarkable application of Campbell's theorem concerning the effect of random events (Campbell and Francis 1946), to deriving the magnitude not only of shot noise in diodes but of thermal noise in ohmic conductors too, was an indication that the two phenomena could not be so very different. More recently still, Fürth (1948) has applied the theory of thermal noise to predicting the shot noise in diodes, even under conditions of space-charge smoothing.

The purpose of this note is to show that sufficient knowledge existed as long as thirty years ago, to allow the prediction of the magnitude of the noise in an ohmic conductor by direct application of Schottky's formula for shot noise.

For practical purposes Schottky's result may be stated as follows: An average direct current I consisting of a random movement in a given direction of charges of magnitude e , carries with it a mean square noise-current in the frequency range f to $f + \Delta f$, given by

$$\overline{(i_f)^2} = 2eI\Delta f. \quad (1)$$

(This holds provided the frequency is not too high, and there is no special smoothing mechanism. Also it is the "short-circuit" noise, *i. e.* the noise when the current flows into a circuit of zero impedance.)

If the terminals of a conductor of resistance R and equilibrium temperature T are short-circuited, any cross-section of the conductor will be crossed by elementary charges (electrons, ions or holes) in their random thermal

* Communicated by the Author.

motion. Some will cross from left to right, say, others from right to left. The former will constitute an average current I_0 and the latter an average current I_0 in the opposite direction. (They must be equal because no nett current flows). Since the random motions which give rise to each of these currents are uncorrelated, the total mean square noise-current arising is equal to the sum of the mean square noise-currents of each.

If now we knew that the elementary charges concerned were all of one kind, and could measure I_0 (*e. g.* by some kind of current rectifier) we could immediately obtain the mean square noise-current by application of equation (1):

$$\overline{(i_f)^2} = 4eI_0 4f. \quad (2)$$

Fortunately this is not necessary, because Einstein's work (1906) on Brownian motion leads to a method of calculating the effective value of eI_0 . Einstein showed that the fluctuations of charge in an electrical circuit can formally be treated in the same way as the fluctuations of displacement of a particle suspended in a liquid, and obtained the result

$$\overline{Q^2} = (2kT/R)t, \quad (3)$$

for the mean square charge crossing any cross-section in a sufficiently long time t . Here k is the Boltzmann constant.

Let us assume that only one type of charge e is present. Then

$$Q = e_1 + e_2 + \dots + e_n,$$

where

$$e_i = \pm e,$$

n is the number of charges that cross in the time t , and e_i is positive or negative, depending on whether the charge concerned crosses from left to right, or right to left.

Therefore

$$Q^2 = \sum e_i^2 + \sum_{i \neq j} e_i e_j,$$

and

$$\begin{aligned} \overline{Q^2} &= \sum \overline{e_i^2} + \sum_{i \neq j} \overline{e_i e_j}, \\ &= ne^2, \quad (4) \end{aligned}$$

since $\overline{e_i e_j} = 0$, positive or negative values being equally likely.

Substituting this result in equation (3) gives

$$(n/t) = (2kT/Re^2). \quad (5)$$

The important expression $2kT/Re^2$ therefore gives the average total number of charges that cross a cross-section per second. Half of these cross from left to right, say, hence

$$I_0 = kT/Re. \quad (6)$$

Substituting this value in equation (2) gives

$$\overline{(i_f)^2} = (4kT/R)\Delta f, \quad (7)$$

which is precisely the ordinary thermal noise result originally discovered and demonstrated in a quite different manner by Nyquist (1928).

The argument still holds if the elementary charges are of more than one type. For example, suppose there are three types of charge magnitudes e , a and o .

Then one will have, similarly,

$$\begin{aligned} Q &= e_1 + \dots + e_n + a_1 + \dots + a_m + o_1 + \dots + o, \\ Q^2 &= \Sigma e_i^2 + \Sigma a_i^2 + \Sigma o_i^2 \\ &\quad + \Sigma \Sigma_{i \neq j} e_i e_j + \Sigma \Sigma_{i \neq j} a_i a_j + \Sigma \Sigma_{i \neq j} o_i o_j \\ &\quad + \Sigma \Sigma e_i a_i + \Sigma \Sigma a_i o_i + \Sigma \Sigma o_i e_i, \end{aligned}$$

so that

$$\overline{Q^2} = ne^2 + ma^2 + lo^2,$$

and hence

$$(n/t)e^2 + (m/t)a^2 + (l/t)o^2 = (2kT/R).$$

Now ne/t , ma/t and lo/t are precisely double the average direct current carried in one direction by each of the three types of charge respectively.

Hence the total $\overline{(i_f)^2}$ is given by

$$\begin{aligned} \overline{(i_f)^2} &= 4e(ne/2t)\Delta f + 4a(ma/2t)\Delta f + 4o(lo/2t)\Delta f \\ &= 2[(ne^2/t) + (ma^2/t) + (lo^2/t)]\Delta f \\ &= (4kT/R)\Delta f. \end{aligned}$$

REFERENCES.

- BULL, C. S., 1947, Discussion on paper by Macdonald, D. K. C., and Fürth, R., *Proc. Phys. Soc.*, **59**, 404.
 CAMPBELL, N. R., and FRANCIS, V. J., 1946, *J.I.E.E.*, **93**, 45.
 EINSTEIN, A., 1906, *Ann. Phys. Lpz.*, **19**, 371.
 FÜRTH, R., 1948, *Proc. Roy. Soc. A*, **192**, 593.
 MOULLIN, E. B., 1938, *Spontaneous Fluctuations of Voltage*, Oxford.
 NORTH, D. O., HARRIS, W. A., and THOMSON, B. J., 1940, *R.C.A. Review*, **3**, 463.
 NYQUIST, H., 1928, *Phys. Rev.*, **32**, 112.
 SCHOTTKY, W., 1918, *Ann. Phys. Lpz.*, **57**, 541.

CXII. *On the Theory of Strength of Quasi-Isotropic Solids.*

By R. FÜRTH,

Birkbeck College, University of London *.

[Received August 29, 1949.]

SUMMARY.

It is shown that Sir Lawrence Bragg's theory of the strength of metals, which is based on the assumption of a block structure of the crystallites, and the author's thermodynamic theory of strength, which is based on the idea of a connection between breaking and melting, can be related to each other if it is assumed that the block structure is an intrinsic feature of the crystal lattice. This can further be linked up with some recent theoretical ideas of M. Born on the one hand, and on the other hand with new experimental observations on the broadening of X-ray lines of metals by W. A. Wood and his school. It appears that the size of the blocks, as determined from these experiments, has the same order of magnitude as the critical block size calculated from Born's theory. Some implications of the concept of an "intrinsic block structure" of crystal lattices are discussed.

1. FROM their measurements of the broadening of X-ray lines, W. A. Wood and collaborators (1939, 1943) have concluded that by progressive deformation the crystal grains of annealed metals are broken up into smaller units which have a minimum size for a given metal. We shall call these units "crystal blocks." Using ideas similar to those expressed by E. Orowan (1941) on the origin of slip bands Sir L. Bragg has suggested a theory of the shear strength of polycrystalline metals (1942, 1948), which relates the size of these crystal blocks (or mosaic fragments, as he calls them) with the bulk strength of the material in a remarkably simple way. The predictions of this theory turned out to be in agreement with the experimental results of Wood and his school. This evidence has been criticized by Lipson and Stokes (1949) on the grounds that the broadening of the X-ray lines could equally well be attributed to inhomogeneous internal strains, but in a recent publication Wood and Rachinger (1948) maintain that this latter explanation of the observed effects has been disproved by their newest measurements. They now give more precise figures for the size of the blocks which are in perfect agreement with the predictions of Bragg's theory.

* Communicated by the Author.

In 1940 I proposed another, seemingly completely different theory of the tensile strength of quasi-isotropic solids, which was based on thermodynamic considerations and led to results agreeing well with experiment. On the other hand M. Born has shown in a recent publication (1947) that the application of quantum theory to the dynamics of crystal lattices leads to difficulties which can be resolved if one assumes a certain limit for the size of a perfect lattice. This can be interpreted to mean that a monocrystal or a crystallite of a polycrystalline material has a kind of "intrinsic block structure." In, so far, unpublished work I have tried to develop a theory of heat conduction in crystals on the basis of this concept and have also applied it to the problem of melting. In the present paper I propose to show that there is in fact a very close connection between Bragg's and my own theory of strength, and that the block structure of metal grains, discovered by Wood, is most probably identical with the intrinsic block structure of Born's theory.

2. I shall first briefly summarize the main argument of my own theory. The principal idea is the assumption of a close relationship between the phenomena of breaking and melting, melting being identical with the breaking up of the solid structure by the thermal movement of the atoms, and vice versa, breaking being nothing else than melting enforced by an applied stress.

Several attempts have been made in the past to calculate the "ideal" tensile strength of a perfect crystal on the basis of the lattice theory. An estimate of the order of magnitude to be expected was first given by M. Polanyi (1921) whose formula is usually quoted in papers and textbooks on this subject. The relation can, however, be expressed in a much more adequate form which can be derived by a simple consideration (Fürth 1940), and which is essentially identical with the result of the rigorous treatment of the problem by lattice theory (Born and Fürth 1940). The maximum stress a perfect lattice can sustain is of the order of magnitude of the latent heat of sublimation per unit volume. It is, however, well known that the actual strength of real solids is, in general, very much smaller than this.

The idea of the connection between breaking and melting suggests that the strength should be related to the heat of melting rather than the heat of sublimation. It can indeed be seen from the available data that the ultimate strength of a quasi-isotropic material is of the same order of magnitude as the latent heat of melting per unit volume, and this can hardly be regarded as a mere coincidence. This can be interpreted by the assumption that the elementary processes which are responsible for the yield and eventually the rupture of a polycrystalline solid material are released by an "activation energy" of the order of magnitude of the latent heat of fusion. This point of view now seems to have been fairly generally accepted (see *e.g.* remarks in recent papers by N. F. Mott (1948) and Sir L. Bragg (1948); its intimate connection with the author's theory is, however, apparently not being realized.

Actually it is possible to treat the process of breaking on a thermodynamical basis without reference to a detailed mechanism at all, just as it is possible to treat the process of melting by the application of thermodynamics to the equilibrium between two phases. The main result of this theory is the following formula relating the tensile strength F of a quasi-isotropic solid at zero temperature with the heat of melting Q , the density ρ and the Poisson ratio μ

$$F = Q\rho \frac{1-2\mu}{3-5\mu}, \quad \dots \dots \dots (1)$$

which I have shown to be in good agreement with experiment*.

3. It is possible to transform the relation (1) into another form in which it can directly be compared with Bragg's relation. For this purpose let us represent the cohesion forces in the solid by the widely used interatomic potential law of Mie (1903) (usually referred to as "Lennard-Jones law")

$$\phi(r) = -ar^{-m} + br^{-n} \quad (m < n), \quad \dots \dots \dots (2)$$

where r is the distance between two atoms, and a , b , m , n are constants. This is justified because it can be shown (Fürth 1944) that the equation of state of solid elements can be well represented by the use of this force law under the assumption that the atoms are closely packed in an arrangement resembling a face-centred cubic lattice. It is further possible to determine in this way from available experimental data the exponents m , n in the force law for most elements with reasonable accuracy.

It follows from the theories of Grüneisen (1912) and Born and Landé (1918) that, provided the force law (2) holds, the following relation exists between the compressibility κ , the density ρ , the sublimation heat Λ , and the constants m , n .

$$\kappa\Lambda\rho = 9/mn. \quad \dots \dots \dots (3)$$

On the other hand one has the well-known relation between the Young modulus E , κ and μ :

$$\kappa = 3(1-2\mu)/E. \quad \dots \dots \dots (4)$$

Thus from (3) and (4)

$$E = \Lambda\rho(1-2\mu)mn/3. \quad \dots \dots \dots (5)$$

Combining (1) and (5) one obtains

$$F/E = 3Q/\Lambda(3-5\mu)mn. \quad \dots \dots \dots (6)$$

* It is well known that by the careful elimination of cracks, especially at the surface, the tensile strength of a specimen can in some cases be increased considerably above the "normal" value so as almost to reach the order of magnitude of the "ideal" strength of the perfect lattice. This fact might, in the framework of this theory, be interpreted as a "superheating" phenomenon.

For a lattice of a given type the Poisson ratio depends on m, n only and can be calculated for various combinations of these constants (Fürth 1942). For the present purpose it is sufficient to compute an average value of the factor $3/(3-5\mu)mn$ appearing in (6). From the table of force constants evaluated by me (Fürth 1944) one obtains the average values $m=4, n=7$ for the majority of the metallic elements. To this corresponds a value $\mu=0.4$ which is indeed close to the average observed value $\mu \simeq 0.36$ of the Poisson ratio. Using these figures one can write (6) in the approximate form

$$F/E \simeq 0.1 Q/A. \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

This relation can be further transformed if one makes use of a relation between Q and the increase ΔV of the specific volume V caused by fusion which was derived by Harasima (1938) from the electron theory of metals and by me (1941 a) from the fundamental concepts of the "hole theory of liquids" (Fürth 1941 b);

$$Q/A = \Delta V/V. \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

This formula was shown to be approximately valid over a wide range of substances. Combining (7) and (8) one obtains

$$F/E \simeq 0.1 \Delta V/V. \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

This relation clearly demonstrates that the process of the loosening of the solid structure under external stress which precedes rupture must be intimately connected with the loosening that is characteristic for the transition from the solid to the liquid state (Fürth 1941 a and c).

It is well known that the terms on both sides of formula (8) are of the same order of magnitude for all elements, and so, according to (7) and (9) the quantity F/E should be approximately constant. This is indeed the case (Fürth 1941 a), and it can be made plausible on theoretical grounds. For F/E is equal to the relative extension of a rod under tensile load at which the process ending in rupture sets in, and it is to be expected from lattice-theoretical considerations (Born and Fürth 1940) that this quantity should have a fixed value.

The average experimental value of the ratio Q/A is

$$Q/A \simeq 1/32 \simeq 3 \times 10^{-2};$$

on the other hand the average experimental value of F/E is (Fürth 1941 a)

$$F/E \simeq 3 \times 10^{-3},$$

so that the relation (7) is indeed satisfied.

4. Bragg's theory of strength, which is based on the assumption of a block structure of the crystallites, leads to the following formula for the shear strength S of a polycrystalline metal:

$$S/G = \alpha d/L, \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

where G is the shear modulus, d the interatomic distance, L the linear dimension of the blocks and α a constant of the order of magnitude unity.

Now suppose that an energy ϵ is needed for removing an atom from its next neighbours in the lattice. Then, if N is the number of atoms per unit mass,

$$A = N\epsilon. \quad (11)$$

But whereas it is necessary to destroy the lattice completely for sublimation to occur, we may assume that melting will take place when the lattice is broken up through a separation of the blocks. Suppose for simplicity the blocks to be cubic in shape. Then, if n is the number of blocks per gram one has evidently

$$n/N = (d/L)^3. \quad (12)$$

Further, as the total number of surface atoms is equal to $6(L/d)^2n$, one has

$$Q \simeq 6n(L/d)^2\epsilon. \quad (13)$$

From (11), (12) and (13) one obtains

$$Q/A \simeq 6d/L. \quad (14)$$

Finally by identifying the ratios F/E and S/G , which one can safely do in a consideration concerning orders of magnitude only, one gets from (7) and (14)

$$S/G \simeq 0.6d/L, \quad (15)$$

which is seen to be practically identical with Bragg's formula (10).

5. It is rather surprising to see that Bragg's theory, for which the assumption of the existence of the lattice blocks is essential, should lead to practically the same result regarding strength as my own theory which makes no such assumption. The explanation is to be sought in the hypothesis formulated above according to which melting consists in the complete break-up of a lattice into separate blocks. Now, as the heat of melting is a fixed quantity for a given substance, not depending on its mechanical treatment, we have to conclude that this block structure must be an *intrinsic* and not a "*structure sensitive*" property of a solid crystalline material, in contrast to the so-called "*mosaic structure*" which can be brought to light by various methods and which strongly depends on the previous mechanical and thermal treatment of the specimen. In other words, we have to assume that the "*intrinsic blocks*" are already formed within the grains of a polycrystalline material, and that the procedure adopted by Wood only serves to make them observable by means of X-rays. This would immediately explain why the minimum size of Wood's crystal units is a fixed quantity for a given substance.

As already mentioned in 1, something like an intrinsic block structure of perfect crystals was suggested by Born (1947) on account of certain difficulties in the application of quantum mechanics to the dynamics of crystal lattices. As a result of a preliminary and rather crude consideration he calculates a certain length l_0 which determines the critical size above which a crystal ceases to be perfect at zero temperature as a result

of the finite amplitude of the zero point vibrations. This l_0 is connected with the atomic distance d , the Debye temperature θ , and the coefficient β of linear thermal expansion at high temperature by the formula

$$l_0 = 2d/\beta\theta. \quad (16)$$

It is interesting to note that the number $z = l_0/d$ of atoms along l_0 is of the same order of magnitude for almost all elements round about $z \simeq 500$.

For finite temperatures T the critical length l decreases according to the formula

$$l = l_0 \tanh \theta/2T. \quad (17)$$

The figures quoted by Wood and Rachinger (1948) give an average

$$L/d \simeq 120.$$

This is of the same order of magnitude as the theoretical value of z , and it strongly suggests that the critical length l of Born's theory is the linear size of actual "intrinsic lattice blocks" within which the lattice is perfect and whose boundaries consist of some intrinsic lattice defects.

A direct comparison between the values of L/d appearing in formula (15) and the value l/d from formula (16) and (17) gives strong support to this view. With the average value $S/G \simeq F/E \simeq 3 \times 10^{-3}$, given above, one obtains from (15)

$$L/d \simeq 200.$$

On the other hand the Debye temperatures for most metals are round about room temperature. Thus one has for such temperatures

$$\tanh \theta/2T \simeq \tanh (1/2) \simeq 0.5,$$

and hence from (17) with $z \simeq 500$

$$l/d \simeq 250,$$

which is indeed close to the value of L/d .

Unless this correspondence is purely accidental, which it is hard to believe, one is driven to the conclusion that the process of melting of solids is intimately bound up with this intrinsic block structure, and that a proper theory of melting has to take this fact into account. It is indeed well known that all theories of melting that are based on the dynamics of perfect lattices, like the theory of Born (1939), are in quantitative disagreement with experiment. For example, it was shown by me (1944) that melting actually occurs with most elements at temperatures about 3.5 times lower than the "critical temperatures" at which the breakdown of the perfect lattice due to thermal vibrations should take place.

It seems very likely that this discrepancy will disappear when proper account of the block structure is taken in the theory. Certain preliminary calculations which I carried out some time ago along these lines show signs of promise, and it is hoped that definite results will be obtained in the not too far future.

REFERENCES.

- BORN, M., 1939, *J. Chem Phys.*, **7**, 591 ; 1947, *Proc. Math. and Phys. Soc. of Egypt*, **3**, 35.
 BORN, M. and FÜRTH, R., 1940, *Proc. Phil. Soc. Cambridge*, **36**, 454.
 BORN, M. and LANDÉ, A., 1918, *Verh. d. Phys. Ges.*, **20**, 210.
 BRAGG, W. L., 1942, *Nature*, **149**, 511 ; 1948, *Phys. Soc.*, *Report of a Conference on Strength of Solids*, Bristol, 1947, 26.
 FÜRTH, R., 1940, *Proc. Roy. Soc. A*, **177**, 217 ; 1940, *Nature*, **145**, 741 ; 1941 a, *Proc. Phil. Soc. Cambridge*, **37**, 34 ; 1941 b, *Ibid.*, **37**, 252 ; 1941 c, *Ibid.*, **37**, 177 ; 1942, *Proc. Roy. Soc. A*, **180**, 285 ; 1944, *Ibid.*, **183**, 87.
 GRÜNEISEN, E., 1912, *Ann Phys.*, **39**, 257.
 HARASIMA, A., 1938, *Proc. Phys. Math. Soc. Japan*, **20**, 850.
 LIPSON, H. and STOKES, A. R., 1949, *Nature*, **163**, 871.
 MIE, G., 1903, *Ann. Phys.*, **11**, 657.
 MOTT, N. F., 1948, *Proc. Phys. Soc.*, **60**, 391.
 OROWAN, E., 1941, *Nature*, **147**, 452.
 POLANYI, M., 1921, *Z. f. Phys.*, **7**, 323.
 WOOD, W. A., 1939, *Proc. Roy. Soc. A*, **172**, 231 ; 1943, *Nature*, **151**, 585.
 WOOD, W. A. and RACHINGER, W. A., 1948, *Nature*, **162**, 891.

CXIII. The Magnetic Double Refraction of Liquid Mixtures.—II. Toluene, and Carbon Disulphide, in Various Solvents, and Relation of MDR to Refractive Index.

By EDWARD J. BURGE, B.Sc. and OLLE SNELLMAN, Fil.dr.,
 Institute of Physical Chemistry, Upsala, Sweden*.

[Received September 12, 1949.]

ABSTRACT.

Measurements of the magnetic double refraction of carbon disulphide in carbon tetrachloride and in cyclohexane are given. Solutions of toluene in carbon tetrachloride, cyclohexane, and heptane are studied with regard to both magnetic double refraction and refractive index. Anomalies observed for the heptane solutions remain in the magnetic measurements after purification of the components, showing that magnetic double refraction is more sensitive than refractive index to structural alterations in the solutions. The refractive index of toluene in cyclohexane does not appear to have been measured previously. Suggestions for further studies are given.

INTRODUCTION.

STUDIES of benzene in various solvents have been reported in a previous paper (Burge and Snellman 1949). The mixtures considered here were chosen, in the first instance, in order to check the results of other investigators and compare the new values with Ramanadham's theory. The apparatus used was the same as that described previously and the symbols used are also the same.

* Communicated by the Authors.

MEASUREMENTS OF OTHER INVESTIGATORS.

Solutions of carbon disulphide in carbon tetrachloride and cyclohexane, respectively, have been studied by Chinchalkar (1933). For the mixture with carbon tetrachloride, three out of four points lie on a smooth curve (fig. 1). Ramanadham (1936) claims fair agreement between his theory and these measurements. Chinchalkar's values for the mixture with cyclohexane lie on a fairly smooth curve (fig. 2).

Fig. 1.

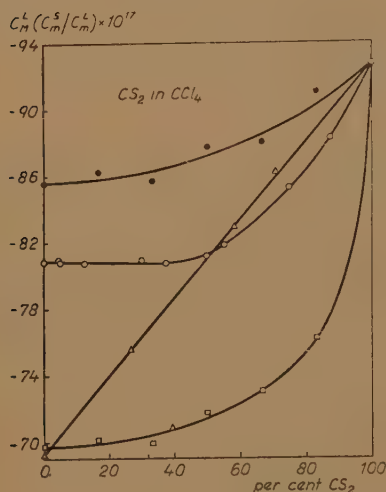
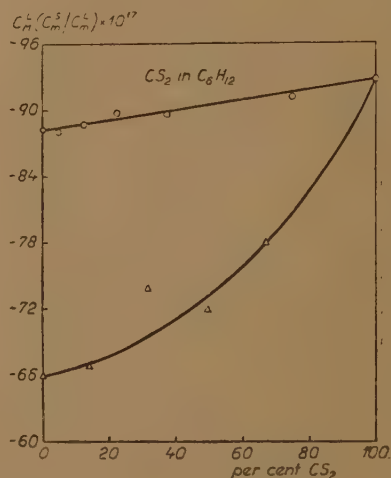


Fig. 2.



- Ramanadham's theory.
- New calculations.
- △ Chinchalkar's measurements.
- New measurements.

The values for toluene in carbon tetrachloride given by Chinchalkar are not very consistent (fig. 3), and the claim of good agreement with theory made by Ramanadham means very little. On the other hand, the values given by Chinchalkar for toluene in cyclohexane lie on a smooth curve (fig. 4). No previous measurement of mixtures of toluene and heptane is reported in the literature.

CARBON DISULPHIDE IN CARBON TETRACHLORIDE.

The sample of carbon disulphide used (Schering-Kahlbaum) had a MDR of $C_m = -6.62 \times 10^{-13}$, using C_m (benzene) $= 7.12 \times 10^{-13}$ (Snellman 1944). This is somewhat larger than the values of Cotton and Mouton, and Chinchalkar, depending on the value assumed for benzene. Its refractive index was 1.6300, compared with the accepted value 1.6295. The MDR of the carbon tetrachloride used was about half of that found for the previous samples, although the refractive index was the same.

The observed values are given in full in Table I. and fig. 1, together with the measurements of Chinchalkar and the theoretical values of Ramanadham. Concerning the measurements it is seen that the values for the higher concentrations agree fairly well, as was found previously.

Fig. 3.

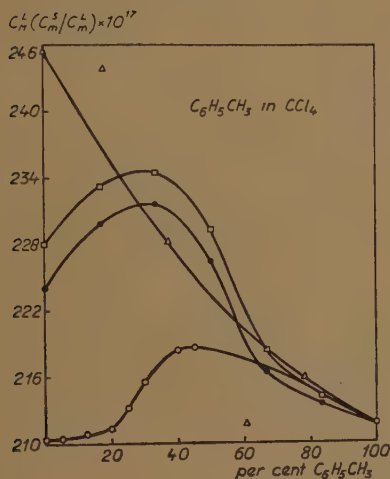
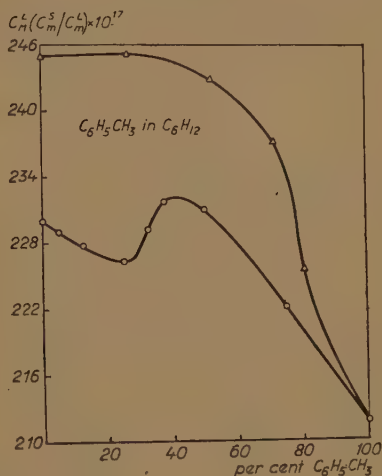


Fig. 4.



- Ramanadham's theory.
 ● New calculations.
 △ Chinchalkar's measurements.
 ○ New measurements.

Three of Chinchalkar's four points lie on a straight line giving $(C_M^S)_0 = -69.2 \times 10^{-17}$, where $(C_M^S)_0$ is the Molecular Cotton-Mouton Constant for a solution, extrapolated to infinite dilution. The new value is $(C_M^S)_0 = -80.82 \times 10^{-17}$.

TABLE I.
Carbon Disulphide in Carbon Tetrachloride.

Com- pensator	CS ₂ %	θ	Δ°	Δ _{LT} °	Δ _C °	$C_M^L(\Delta_C/\Delta_{LT}) \times 10^{17}$
B ₁	0	— 0 12	— 0.06	—	—	—80.82
B ₁	4.65	— 2 46	— 0.83	— 0.885	— 0.773	—80.96
B ₁	5.00	— 2 57	— 0.885	— 0.95	— 0.828	—80.78
B ₁	12.5	— 7 4	— 2.12	— 2.375	— 2.07	—80.77
B ₂	30.0	— 6 58	— 5.017	— 5.70	— 4.975	—80.89
B ₂	37.5	— 8 42	— 6.23	— 7.13	— 6.21	—80.72
B ₂	50.0	—11 50	— 8.35	— 9.50	— 8.32	—81.17
B ₂	55.56	—13 22	— 9.35	—10.556	— 9.32	—81.84
S	75.0	—141	—13.13	—14.25	—13.11	—85.27
S	87.5	—170	—15.84	—16.625	—15.83	—88.25
S	100.0	—204	—19.00	—19.00	—19.00	—92.68

The theoretical curve of Ramanadham has a shape somewhat similar to that for the present observations, but requires a large change of C_M^S between 100 per cent and 70 per cent CS_2 . The unusual magnitude of this change suggested that a check should be made of Ramanadham's calculations. The new values were found to be very different from the old (Table II. and fig. 1). The shape is still the same but the curve does

TABLE II.

Carbon Disulphide in Carbon Tetrachloride.

Ramanadham's Theory.

$$\begin{array}{ll}
 n_1 = 1.6222 & (C_M^L)_{CS_2} = -92.68 \times 10^{-17} \\
 n_1/(n_1^2 - 1) = 0.9943 & n_{ccl_4} = 1.4562 \\
 B_1^* = 20.540 \times 10^{-24} & b' = 1.040 \times 10^{-23} \\
 B_2^* = 9.477 \times 10^{-24} & \nu_{ccl_4} = 6.241 \times 10^{21} \\
 b_1 = 14.340 \times 10^{-24} & \nu_{CS_2} = 9.99 \times 10^{21} \\
 b_2 = 5.870 \times 10^{-24} &
 \end{array}$$

$x\%CS_2$	$(n^2-1)/n$	$\frac{B_1-B_2}{B_1^*-B_2^*}$	C_m^S/C_m^L	$\frac{C_M^L(C_m^S/C_m^L)(x/100) \times 10^{17}}{\text{Ramanadham} \quad \text{New values}}$	
16.7	0.8072	0.9166	0.155	-70.1	-86.19
33.3	0.8446	0.9141	0.308	-69.9	-85.64
50.0	0.8815	0.9434	0.473	-71.7	-87.68
66.7	0.9221	0.9486	0.633	-73.0	-87.95
83.3	0.9639	0.9806	0.818	-76.2	-91.01

not agree with the new measurements, and it must be concluded that Ramanadham's theory does not hold for this case. It can be seen that the claim of fair agreement made by Ramanadham is not justified—only one point of Chinchalkar's observations lies on the curve, and it does not agree with the other three points. The value of $(C_M^S)_0$ given by the new calculations is -85.54×10^{-17} , compared with the value -69.7×10^{-17} given by Ramanadham.

CARBON DISULPHIDE IN CYCLOHEXANE.

The cyclo-hexane used (Eastman Kodak) had a refractive index of 1.4275 (accepted value 1.4290), and $C_m = -1.31 \times 10^{-14}$. The values obtained are given in condensed form in Table III., and plotted, together

TABLE III.

Carbon Disulphide in Cyclohexane.

$C_M^L(C_m^S/C_m^L) \times 10^{17}$	$CS_2\%$	0	4.75	12.5	25.0	37.5	75.0
		-88.2	-88.05	-88.69	-89.81	-89.62	-91.10

with those of Chinchalkar, in fig. 2. There is not even agreement with the values at the higher concentrations, the discrepancy being in the same direction as for the mixtures with carbon tetrachloride. The value of $(C_M^S)_0$ is -88.2×10^{-17} , which is to be compared with -65.99×10^{-17} given by Chinchalkar.

DISCUSSION OF RESULTS FOR CARBON DISULPHIDE.

The variation of C_M^s with concentration is found to be less than that given by Chinchalkar, but greater than that expected on Ramanadham's theory, according to the new calculations. It appears that the theory cannot be applied to an asymmetric molecule such as carbon disulphide, this being borne out by the fact that it does not hold for nitrobenzene, or toluene.

TOLUENE IN CARBON TETRACHLORIDE.

The sample of toluene used (Schering, $n=1.4971$ —accepted value 1.4978) had a MDR of $C_m=7.64 \times 10^{-13}$. This is to be compared with 7.30×10^{-13} (Snellman). The carbon tetrachloride (Schering, $n=1.4620$) had the same MDR as that used for the mixtures with benzene.

The observed values lie on a smooth curve (fig. 3), the shape of which agrees with the theoretical curve of Ramanadham. However, the values (Table IV.) do not agree in order of magnitude, presumably due to the fact that toluene, like carbon disulphide, is an asymmetric molecule.

TABLE IV.

Toluene Mixtures.

Carbon tetrachloride		Cyclohexane	
$C_6H_5CH_3$ %	$C_M^L(C_M^S/C_M^L) \times 10^{17}$	$C_6H_5CH_3$ %	$C_M^L(C_M^S/C_M^L) \times 10^{17}$
0	210.4	0	230.0
5.0	210.4	4.76	229.0
12.5	210.8	12.5	227.8
20.0	211.3	25.0	226.3
25.0	213.2	32.5	229.2
30.0	215.5	37.5	231.8
40.0	218.4	50.0	231.0
45.0	218.6	75.0	222.1
100.0	211.7	100.0	211.7

A recalculation of Ramanadham's theoretical values gives almost the same points (Table V). The difference is almost certainly due to the fact that different constants have been used in the calculation, since b_1 and b_2 for toluene are not given by Ramanadham (1934). They are taken from Stuart (1934). It is seen that the measurements of Chinchalkar are rather scattered, and a possible curve gives $(C_M^S)_0=245 \times 10^{-17}$. The values from Ramanadham's theory are: Ramanadham, 228×10^{-17} , new calculations, 224×10^{-17} . The present measurements give $(C_M^S)_0=210 \times 10^{-17}$, following a maximum of 218.6×10^{-17} at about 45 per cent toluene.

TABLE V.

Toluene in Carbon Tetrachloride.

Ramanadham's Theory.

$$\begin{aligned}
 n_1 &= 1.4917 & (C_M^L)_{C_6H_5CH_3} &= 211.7 \times 10^{-17} \\
 n_1/(n_1^2 - 1) &= 1.2175 & n_{ccl_4} &= 1.4554 \\
 B_1^* = b_1(1 + p_1\chi) &= 12.286 \times 10^{-24} & b' &= 1.040 \times 10^{-23} \\
 B_2^* = b_2(1 + p_2\chi) &= 18.917 \times 10^{-24} & \nu_{ccl_4} &= 6.241 \times 10^{21} \\
 b_1 &= 7.48 \times 10^{-24} & \nu_{C_6H_5CH_3} &= 5.6584 \times 10^{21} \\
 b_2 &= (13.66 + 15.64)/2 = 14.65 \times 10^{-24} & p_1/4\pi &= .5244 \\
 \chi &= (n_1^2 - 1)/4\pi & p_2/4\pi &= .2378
 \end{aligned}
 \left. \begin{array}{l} \\ \\ \\ \\ \end{array} \right\} \begin{array}{l} \text{Extra-} \\ \text{polated} \\ \text{values.} \end{array}$$

$x\% C_6H_5CH_3 \quad (n^2 - 1)/n$	$\frac{B_1 - B_2}{B_1^* - B_2^*}$	C_M^S/C_M^L	$\frac{C_M^L(C_M^S/C_M^L)(x/100) \times 10^{17}}{\text{Ramanadham} \quad \text{New values}}$	
			Ramanadham	New values
16.7	0.7782	1.0843	233.2	229.8
33.3	0.7880	1.0902	234.5	231.6
50.0	0.7972	1.0668	229.3	226.4
66.7	0.8048	1.0205	218.3	216.3
83.3	0.8144	1.0057	214.2	213.5

TOLUENE IN CYCLOHEXANE.

The same sample of cyclohexane was used as for the measurements with carbon disulphide. The toluene, however, was from a different sample (Schering, $n=1.4972$) with $C_m=7.70 \times 10^{-13}$.

Whereas Chinchalkar's values lie on a smooth curve giving $(C_M^S)_0=245 \times 10^{-17}$, the new values show a maximum value of 232.2×10^{-17} for about 40 per cent toluene, falling to 226.2×10^{-17} at 23 per cent and rising again to $(C_M^S)_0=230 \times 10^{-17}$. The results are given in Table IV. and fig. 4.

REFRACTIVE INDEX OF TOLUENE IN CARBON TETRACHLORIDE,
AND IN CYCLOHEXANE.

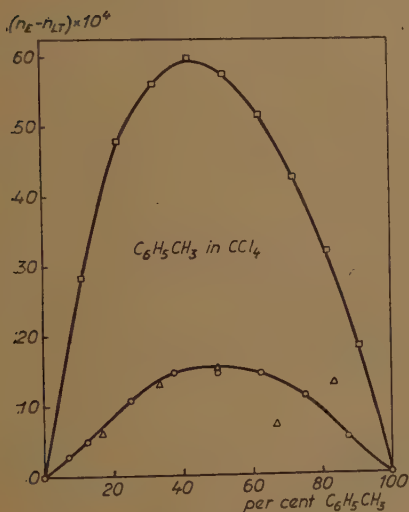
Ramanadham's theory uses experimental values of the refractive indices of the mixtures considered. The value of C_M^S by this theory has a maximum value for 30 per cent toluene, and from the new measurements passes through a maximum at about 45 per cent. It appeared, therefore, that this anomaly may be detectable in the values of the refractive index. The results of previous investigators were compared with measurements on the solutions used here, and are presented in fig. 5. The values are all reduced to volume per cent, and the graph gives the divergence from a linear dependence of refractive index on concentration.

If n_1, n_2 are the refractive indices of the solute and solvent, respectively, and x is the percentage of solute, then

$$n_{LT} = n_2 + (n_1 - n_2)x/100. \quad (1)$$

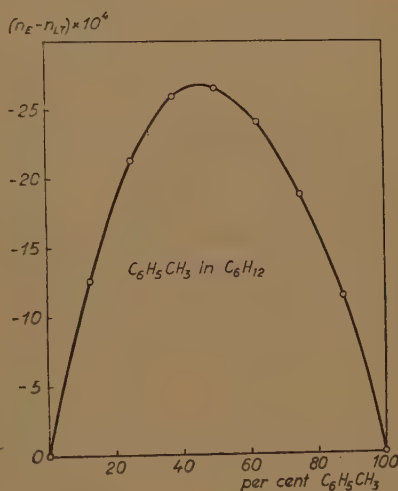
The ordinate is $n_E - n_{LT}$, where n_E is the value from experiment and n_{LT} is from this linear theory. The values of Ramanadham (1936) show an anomaly for concentrations greater than 50 per cent toluene, but the new values (Table VI.) lie on a smooth curve and agree with Ramanadham's below 50 per cent. This is, however, fortuitous since they are for a different wavelength and temperature. The absolute values do not agree, but we are here only concerned with a possible anomalous dependence on concentration. Lehfeldt (1898) gives values that lie on a smooth curve but depart much more from the linear theory than the other values. His values for the pure substances agree well with the present values but not with the accepted values. The measure-

Fig. 5.



- Lehfeldt.
 △ Ramanadham.
 ○ New measurements.

Fig. 6.



- Authors.

ments of Krchma and Williams (1927) give negative values for $n_E - n_{LT}$ and absolute values that are very different from those of other workers.

It may be concluded that the anomaly in the dependence of MDR on concentration for toluene in carbon tetrachloride does not appear in the refractive index.

No previous measurements of the refractive index of toluene in cyclohexane could be found in the literature with which to compare the present values (Table VI.). It is to be noticed that $n_E - n_{LT}$ is negative. As seen from fig. 6, the curve is quite smooth and shows no sign of the anomaly detected in the MDR. These results are striking, since they emphasize the sensitive dependence of MDR on liquid structure, and the relatively less sensitive dependence of refractive index.

TABLE VI.
Toluene Mixtures.

$C_6H_5CH_3\%$	Carbon Tetrachloride		Cyclohexane	
	n_E	$(n_E - n_{IT}) \times 10^4$	n_E	$(n_E - n_{IT}) \times 10^4$
0	1.4610	0	1.4265	0
6.98	1.4638	+ 2.8	—	—
12.5	1.4660	+ 4.9	1.4341	-12.6
25.0	1.4711	+10.7	1.4421	-21.3
37.5	1.4760	+14.6	1.4505	-25.9
50.0	1.4805	+14.5	1.4593	-26.5
62.5	1.4850	+14.4	1.4684	-24.1
75.0	1.4892	+11.2	1.4778	-18.8
87.5	1.4931	+ 5.1	1.4874	-11.4
100.0	1.4971	0	1.4974	0

TOLUENE IN HEPTANE.

The toluene used for the first measurements with heptane was the same as that used for the mixtures with cyclohexane. The heptane (May and Baker, $C_m = -1.01 \times 10^{-14}$) had a refractive index of 1.4006 which is very different from the value ($n = 1.3877$) given in the literature for the normal isomer. However, since the MDR was in good agreement with the values of Schérer (1934), it was expected that the measurements with toluene would give the correct variation of C_M^S with concentration. As a result of these investigations this assumption appears to be unjustified.

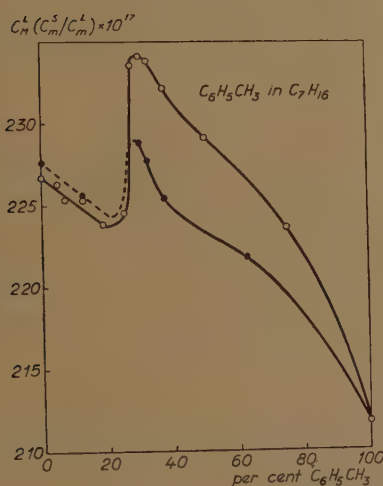
TABLE VII.
Toluene in Heptane.

$C_6H_5CH_3\%$	$C_M^L(C_m^S/C_m^L) \times 10^{17}$	
	Distilled	Undistilled
0	227.6	226.7
12.5	225.6	225.3
30.0	228.8	234.1
32.5	227.7	233.8
37.5	225.4	232.1
62.5	221.8	—
75.0	—	223.6
100.0	211.7	211.7

The measurements are given in fig. 7 and Table VII., and it is seen that there is a marked maximum of C_M^S at 30 per cent toluene. An investigation of the refractive index of the same mixtures also revealed an anomaly at 30 per cent (Table VIII., and fig. 8). The measurements of Bromiley and Quiggle (1933) suggest a possible departure from a smooth curve, but this may be due to experimental errors. Briegleb (1932) gives values

for mixtures of heptane and various solutes, including toluene, but fails to give the refractive index of heptane. The average of the extrapolated values for nine solutes is $n=1.4024$ (Briegleb 1931), which is even larger than n for the sample used here. From $n=1.4024$, positive values are obtained for n_E-n_{LT} (maximum about $+7 \times 10^{-4}$ at 50 per cent),

Fig. 7.



Δ Bromiley and Quiggle.
○ Before distillation.
● After distillation.

Fig. 8.

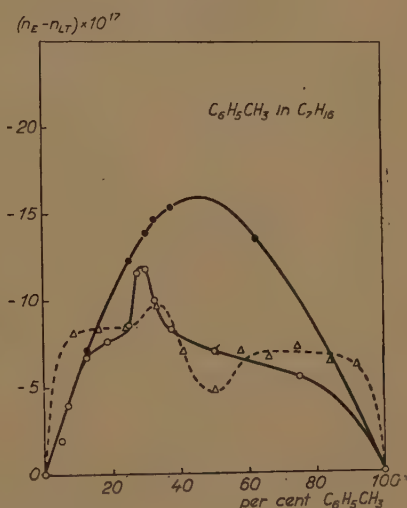


TABLE VIII.

Toluene in Heptane.

$C_6H_5CH_3\%$	Distilled		Undistilled	
	n_E	$(n_E - n_{LT}) \times 10^4$	n_E	$(n_E - n_{LT}) \times 10^4$
0	1.4018	0	1.4006	0
12.5	1.4130	- 7.1	1.4120	- 6.7
25.0	1.4244	-12.3	1.4239	- 8.5
30.0	1.4290	-13.9	1.4284	-11.8
32.5	1.4313	-14.7	1.4310	-10.0
37.5	1.4360	-15.4	1.4360	- 8.3
62.5	1.4600	-13.6	—	—
75.0	—	—	1.4725	- 5.5
100.0	1.4971	0	1.4972	0

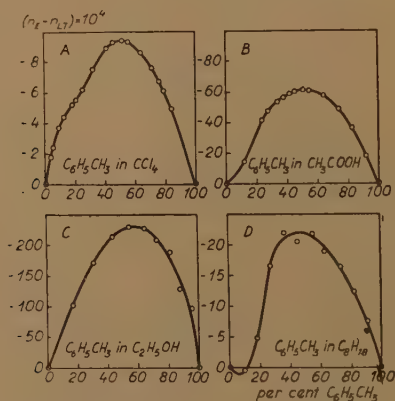
whereas those plotted in fig. 8 are negative. The only other measurements are by Dobrosserdow (1912), and give large positive values for n_E-n_{LT} (maximum $+28 \times 10^{-4}$ at 50 per cent), and $n=1.4066$, and $n=1.4926$ for heptane and toluene respectively. These values are in considerable disagreement with those of other workers.

The high value for n is probably not caused by a many-branched isomer since none of those given in the literature has a value greater than $n=1.3937$. A purification and distillation, following the method given by Weissberger and Proskauer (1935), gave fractions with n varying from 1.3966 to 1.4021 for boiling points from 93°C . to 98°C ., and a residue with $n=1.4052$. The accepted boiling point is 98.4°C ., and for this reason the distillate collected between 97°C . and 98°C . (about 70 per cent of the whole) was used for further measurements. The sample gave $n=1.4018$, and $C_m = -1.10 \times 10^{-14}$. The toluene was also purified and distilled, and successive fractions obtained at 110°C . gave the same refractive index, $n=1.4971$, the MDR being unaltered. The new values for MDR are given in Table VII. and fig. 7, and, although incomplete, suggest a maximum at 30 per cent toluene, but not so pronounced as for the undistilled components. However, the refractive index of these solutions (Table VIII. and fig. 8) shows no sign of an anomaly at this concentration. It is clear that the MDR is sensitive to some change in the liquid that does not affect the refractive index. It is not possible to say whether the anomaly would disappear completely for really pure solutions. The value $(C_M^S)_0$ does not seem to depend appreciably on the purity, and may be given as 227×10^{-17} .

REFRACTIVE INDICES OF MIXTURES OF TOLUENE AND VARIOUS SOLVENTS.

The measurements of toluene and chloroform by Rosanoff, Bacon and White (1914) appear to be very reliable and lie on a smooth curve (fig. 9 A).

Fig. 9.



A Rosanoff, Bacon and White.

B Zawidzki.

C Lehfeldt.

D Kowalski and Modzelewski.

The curve of $n_E - n_{LT}$ against volume per cent is not symmetrical and suggests that measurements of MDR may be of interest. The maximum value of $n_E - n_{LT}$ is very small (9.4×10^{-4}). Toluene in acetic acid has been studied by Zawidzki (1900) and the values lie on a smooth curve

(fig. 9B) which is symmetrical. Mixtures with ethyl alcohol yield a curve that is fairly smooth (fig. 9C) as given by the measurements of Lehfeldt (1898). The maximum of the curve occurs at 58 per cent, and the scatter above 73 per cent may be due to experimental error. The departure from a linear theory is very marked. The measurements of Kowalski and Modzelewski (1901) on toluene and ethyl alcohol do not yield satisfactory points, although given to five places of decimals. The values for N-octane are given by Bromiley and Quiggle (1933) and when plotted as in fig. 9D yield a curve that shows some anomalies and suggests that MDR measurements would be of interest.

GENERAL DISCUSSION.

Of all the substances investigated only toluene has a dipole moment. Although it is small it may play a rôle in the explanation of the observed anomalies. Briegleb has determined the molecular Kerr constant in a series of toluene-heptane mixtures, but only for molar fractions less than 0.45. Throughout this range the constant decreased with decreasing concentration of toluene. For the highest concentration measured the Kerr constant was higher than for pure toluene, and must therefore have a maximum value for a certain concentration. Briegleb has assumed that some of the molecules in toluene are coupled with their CH_3 groups parallel on account of their dipole moments. If the molecules are arranged in such a manner, the anisotropy of the coupled molecules is greater and they will give a larger effect than the single molecules. As the toluene concentration is diminished, the coupled molecules begin to dissociate and the double refraction ought to decrease. This explanation would appear to apply to the results reported here. The small concentration range over which the decrease takes place for cyclohexane and heptane is rather surprising.

SUGGESTIONS FOR FURTHER WORK.

(i) The liquids used, both as solvents and solutes, should be freshly distilled and as pure and dry as possible.

(ii) The refractive index of each mixture should be determined to at least four places of decimals and the values plotted in the form $n_E - n_{LT}$ versus concentration for the detection of possible anomalies.

(iii) The temperature should be controlled, and, if possible, the variations of MDR and refractive index with temperature should be measured.

(iv) It would be of interest to continue the study of mixtures with hydrocarbons, possibly with olefines as well as paraffins. This would provide information regarding the liquid structure of long straight-chain hydrocarbons.

(v) The values of the MDR of vapours are only available for benzene and nitrobenzene. Further values are needed for comparison with theoretical values for the pure liquids, and at infinite dilution. (Snellman 1949).

ACKNOWLEDGEMENTS.

One of us (E.J.B.) gratefully acknowledges the financial support of the Swedish Government and the Swedish Institute, and the fine facilities made available by Professor The Svedberg. Thanks are also due to Mr. Evald Hellman for help with the calculations.

REFERENCES.

- BRIEGLEB, G., 1931, *Z. phys. Chem.*, **14B**, 97; 1932, *ibid.*, **16B**, 249.
 BROMILEY, E. C., and QUIGGLE, D., 1933, *J. Ind. Eng. Chem.*, **25**, 1136.
 BURGE, E. J., and SNELLMAN, O., 1949, *Phil. Mag.*, **40**, 000.
 CHINCHALKAR, S. W., 1933, *Ind. J. Phys.*, **7**, 491.
 DOBROSSERDOW, D., 1912, *J. Soc. Phys. Chem. Russe*, **44**, 396. (In Russian.)
 KOWALSKI, J., and MODZELEWSKI J., 1901, *C. R.*, **133**, 33.
 KRCHMA, J., and WILLIAMS, J. W., 1927, *J. Am. Chem. Soc.*, **49**, 2408.
 LEHFELDT, R. A., 1898, *Phil. Mag.* V, **46**, 42.
 RAMANADHAM, M., 1934, *Proc. Ind. Acad. Sci.*, **1**, 281; 1936, *ibid.*, **3A**, 384.
 ROSANOFF, M. A., BACON, C. W., and WHITE, R. M., 1914, *J. Am. Chem. Soc.*, **36**, 1803.
 SCHÉREER, M., 1934, *Thèse* (Paris).
 SNELLMAN, O., 1944, *Dissertation* (Uppsala); 1949, *Phil. Mag.*, **40**, 000.
 STUART, H. A., 1934, *Molekülstruktur* (Berlin).
 WEISSBERGER, A., and PROSKAUER, E., 1935, *Organic Solvents* (Oxford).
 ZAWIDZKI, J., 1900, *Z. phys. Chem.*, **35**, 129.

CXIV. On the so-called "Clock-paradox" of Special Relativity.

By Professor E. A. MILNE F.R.S. and G. J. WHITROW*.

[Received September 14, 1949.]

ABSTRACT.

It is shown that the "clock-paradox" of special relativity (according to which of two observers who part company, travel with a large relative speed and re-join one another, one will record the lapse of time as shorter than the other) is resolved by kinematic relativity.

1. FORMULATION OF THE PARADOX.

REFERENCE is often made, both in popular writings and in scientific papers, to the so-called "clock-paradox" of special relativity. According to this, if two observers A and B part company from one another and B, after cruising through space with a velocity relative to A which is a considerable fraction of the speed of light, subsequently returns to coincidence with A, then the time that has elapsed in B's experience will be much shorter than the time interval experienced by A. This is supposed to

* Communicated by the Authors.

be a simple, if striking, consequence of the result in special relativity that when A and B are in relative motion in a straight line with speed V each will regard the other's clock as running slow, according to the formula

$$\Delta t'_B = \Delta t_B (1 - V^2/c^2)^{1/2},$$

where t'_B is an epoch of an event at B as measured by B's clock, t_B the epoch of the same event as reckoned by A's clock according to A's convention concerning the epoch of occurrence of a distant event. As long as A and B are in uniform relative motion with speed V , this appears to be compatible with the relation

$$\Delta t_A = \Delta t'_A (1 - V^2/c^2)^{1/2},$$

where t_A is the epoch of an event at A as measured by A's clock, and t'_A the epoch of the same event as reckoned by B.

The paradox arises from the consideration that apparently B could equally have regarded A as cruising through space with relative speed V , in which case it should be the lapse of time in A's experience that is much shorter than that experienced by B.

2. PREVIOUS DISCUSSIONS OF THE PARADOX.

Before examining whether the paradox is real or apparent, it is of interest to summarize published views on the subject. The paradox was first enunciated by Einstein himself in 1905 (Einstein 1905). An account of the paradox and its resolution by means of considerations arising from general relativity has been given by R. C. Tolman (1934), and more recently by E. L. Hill (1947). The youthfulness of cruising observer B as compared with stay-at-home A has recently been described by Lord Russell (1948) who attributes one of his illustrations to Reichenbach. "If two pieces of matter (say the earth and a comet)," he argues, "meet and part and meet again, and if in the interval their relative velocity has been very great, the physicists (if any) who live on the two pieces of matter will form different estimates of the lapse of time between the two meetings." It need hardly be emphasized that such a conclusion does not necessarily follow from the special theory of relativity, and we shall, in fact, show that, if two *equivalent* observers (*i. e.* two observers who stand in a symmetrical relation to one another) part company and move in any way, preserving their relation of symmetry, whenever they meet again their clock-readings will necessarily agree.

It is generally assumed that the paradox arises ultimately from the circumstance that special relativity is concerned only with relative *velocities* of observers, and does not consider the effects of relative *accelerations*; and that, therefore, reference must be made to general relativity as in Tolman's discussion. Alternatively, the ideas of special relativity must be extended to include the effects of *any* kind of relative

motion of observers. Such an extension is provided by kinematic relativity, and it is the object of the present note to show how this discipline provides a sufficient *and necessary* technique for resolving all forms of the paradox.

3. MATERIAL CLOCKS AND IDEAL TIME-SCALES.

Before considering the paradox proper, two distinct problems must be distinguished. The first is the problem of the actual behaviour of material time-keepers when they are subjected to given motions and changes of motion. The second is that of the congruence or non-congruence of the ideal time-scales adopted by observers. Previous to about 1936 it was assumed without question that there is only one ideal time-scale prescribed by the laws of physics and that this is the standard for all "good" natural time-keepers. However, just as the discovery of non-Euclidean geometries suggested that different geometries may be appropriate in different contexts, so recent analysis of the nature of time-keeping suggests that different time-scales may be appropriate in different contexts. This idea was partly anticipated by P. W. Bridgman (1927): "It has always been very puzzling to understand why Einstein has so strenuously insisted that the shift towards the infra-red is an integral part of the general theory, and that, if the shift is not found, the theory must fall. In other words Einstein insists that the assumption that an atom is a clock is an integral part of his theory... Since Einstein created the theory of Relativity it is perhaps ungracious to question his right to stipulate that 'the atom is a clock' is an integral part of the theory. This however degenerates to a mere matter of language, and does not touch the arbitrary nature of the procedure. It does not prevent us from having a second brand of relativity, that of X instead of Einstein, exactly like that of Einstein except that perhaps now the 'clock' is constructed in terms of the life-period of a radioactive disintegrating element." In Bridgman's opinion, "The only way to eliminate the arbitrariness seems to be to postulate that *all* natural processes which run naturally of themselves independently of what we may do may equally well serve as clocks and give the same results."

There is, however, no reason for supposing this to be true, and recent researches suggest the contrary. In another field, Alexis Carrel (1948) suggested that "In short, time is a specific character of things. Its nature varies according to the constitution of each object. Human beings have acquired the habit of identifying their duration, and that of all other beings, with the time shown by clocks. Nevertheless our inner time is as distinct from, and independent of, this extrinsic time as our body is, in space, distinct from, and independent of, the earth and the sun.... When infancy and old age are expressed in solar years, infancy appears to be very short and old age very long. On the contrary, measured in units of physiological time, infancy is very long and old age very short.... We have mentioned that physiological time is quite different from

physical time." Again, with reference to the ageing of the organism, "the number of units of physical time corresponding to a unit of physiological time becomes progressively greater."

The temporal behaviour of a given mechanism, whether physical or physiological, when transported in any manner cannot be prescribed *a priori*. All a general abstract theory can do is to correlate the time-scales used by different observers. Such a correlation can be made to depend on the state of relative motion of the observers if we introduce the notion of equivalence, *i. e.* congruence or symmetry, of the observers. In this context, the clock-paradox does not arise.

4. RESOLUTION OF THE PARADOX IN KINEMATIC RELATIVITY.

It has been shown by the authors that the clock-paradox can be resolved by kinematic relativity. As the proof is contained in a journal (Milne and Whitrow 1938) which is not easily accessible, we reproduce it here.

Let A, B be two observers, carrying congruent clocks, who part company at epoch t_0 by A's clock. Then if $\theta(t)$ be the signal function connecting A and B, $\theta(t)$ will be the time according to B at which he receives a signal emitted by A at time t according to A, and reciprocally. Hence, according to B, the epoch at which they part company must be $\theta(t_0)$. Now, if a signal be emitted by A at epoch t_0 by A's clock and be instantaneously reflected on arrival at B, it returns to A at epoch $\theta\theta(t_0)$ according to A. Since A and B coincide at t_0 , it follows that t_0 must be a root of the equation

$$\theta\theta(t_0)=t_0.$$

We now prove that $\theta(t_0)=t_0$. For, suppose that $\theta(t_0)>t_0$, then since $\theta(t)$ is a monotonic increasing function (Whitrow 1936), it would follow that

$$\theta\theta(t_0)>\theta(t_0)>t_0.$$

Similarly, if $\theta(t_0)<t_0$, then

$$\theta\theta(t_0)<\theta(t_0)<t_0.$$

Hence we must have $\theta(t_0)=t_0$, and so the clocks kept by two equivalent observers necessarily agree at the epoch when they part company, and, moreover, by exactly the same argument, we see that this must also be the case at any subsequent epoch of coincidence should they meet again. This holds irrespective of whatever path the observers describe relative to each other.

The proof is based on the idea that, throughout the relative motion of A and B, these observers are supposed to remain equivalent, and therefore to stand in a symmetrical relation to one another. Their clocks then remain always congruent and agree whenever they coincide. It should be noted that we have not assumed that A and B must be at relative rest at either the beginning or the end of the relative journey.

We have, however, implied that, if say B is accelerated relative to a Galilean frame in one direction, at any point of the journey, then A is similarly accelerated in the opposite direction.

It should be borne in mind that the classical Lorentz formulæ are valid only for a pair of observers who are symmetrical.

5. THE PARADOX IN SPHERICAL SPACE.

We have already remarked that it is generally assumed: (i) that the paradox originally arose from the circumstance that special relativity is concerned with the relative velocities of observers, and has nothing to say about the effects of accelerations, and (ii) that the paradox can be resolved by appeal to general relativity. It is possible, however, to formulate the paradox in a context which involves no reference whatsoever to accelerations. In this case we can no longer apply the technique of general relativity, and the paradox can now be resolved *only* by means of kinematic relativity.

Let us consider observers in a space of constant positive curvature. Let A be an observer who considers himself as at rest, and B a second observer who parts company from A and describes a geodesic relative to A with uniform speed. When B returns to A from the opposite side, B's clock will have run slow as compared with A's. (This can be established by simple light-signalling between A and B). If $t=0$ and $t'=0$ when B leaves A, and if $t=T$ by A's clock when B returns to A, then B's clock will record a time $t'=T(1-V^2/c^2)^{1/2}$ at this event. Were A and B on an equal footing, we could equally well supposed B to be "at rest" and A to have been in motion along the geodesic through B. Hence a contradiction would arise. As no accelerations are involved at any stage, it is clear that the contradiction can only be avoided by assuming that A and B are not on equal footing and that the only observers equivalent to A in this spherical space are those which are fixed relative to A, *i. e.* the corpus of relative stationary observers comprises the totality of observers equivalent to A. By regarding Euclidean space as the limiting form of either spherical or hyperbolic space (of constant curvature), and making use of an obvious continuity postulate, we deduce that in Euclidean and hyperbolic spaces no moving observers are equivalent to the corpus of observers who are "fixed" in the space, *i. e.* who are relatively stationary.

This is the situation in τ -time in the relatively stationary substratum of kinematic relativity. Regrading each observer's clock in the usual way we arrive at the substratum in t -time, in which the only congruent clocks are those which diverged from coincidence with one another at the singular epoch $t=0$. The equivalence of *all* frames in uniform relative motion which is postulated in special relativity and leads to the clock-paradox is now replaced by the equivalence of the more restricted class of frames which coincided at $t=0$.

The present discussion thus leads to the far-reaching conclusion that *the restriction of equivalence to a triply infinite system of fundamental*

observers is essential if paradoxes are to be avoided, and is not merely the consequence of an arbitrary definition (Page 1936). Kinematic relativity provides both a sufficient and a necessary technique for the complete resolution of all forms of the clock-paradox. We consider that this constitutes a new and powerful argument in its favour.

6. PHOTONS AND PROPER-TIME.

In conclusion we consider briefly the application of these ideas to photons. It is well known that the proper-time associated with a typical photon is zero so that, if we could associate a clock with a light-corpuscle of speed c , the interval of time recorded by such a clock while the photon traverses the path would be zero. This situation would appear to be the limiting form of that discussed in the usual formulation of the clock-paradox, but on the present view there is a certain dichotomy between the photon case and that of any observer moving with a relative speed close to c . So far, however, from being a drawback to the present theory, we claim that it is a definite advantage because a clear-cut division must be made between a (material) particle carrying an "observer" moving with any relative speed less than c and a light corpuscle moving with speed c . In relativistic theory, the speed of a material massive particle can no more attain the value c than, in classical theory, it can attain an infinite value. Moreover, a particle moving in a straight line with speed c cannot carry a clock which can be correlated with other clocks. There is, thus, in this sense, a "discontinuity" between the case of the particle of speed less than c and that of the photon of speed c . No hypothetical observer moving with speed c could be regarded as equivalent to any observer moving with lower speed, just as no particle of speed c can be regarded as having a rest-mass equal to that of a massive particle moving with lower speed.

REFERENCES.

- BRIDGMAN, P. W., 1927, *The Logic of Modern Physics* (New York), p. 176.
 CARREL, A., 1948, *Man the Unknown* (London), p. 156 *et seq.*
 EINSTEIN, A., 1905, *Ann. d. Phys.*, (4), **17**, 891; 1911, *Ibid.*, (4), **36**, 898.
 HILL, E. L., 1947, *Phys. Rev.*, **72**, 236.
 MILNE, E. A., and WHITROW, G. J., 1938, *Z. f. Astrophys.*, **16**, 352.
 PAGE, L., 1936, *Phys. Rev.*, **49**, 254.
 RUSSELL, B. A. W., 1948, *Human Knowledge* (London), p. 291.
 TOLMAN, R. C., 1934, *Relativity, Thermodynamics and Cosmology* (Oxford: University Press), p. 194.
 WHITROW, G. J., 1936, *Proc. Lond. Math. Soc.* (2), **41**, 421.

CXV. *The Decay of μ -Mesons.*

By J. H. DAVIES, W. O. LOCK and H. MUIRHEAD,
The H. H. Wills Physical Laboratory, University of Bristol*.

[Received September 30, 1949.]

ABSTRACT.

The energy of the charged particles from the decay of μ -mesons in photographic emulsions has been determined, using multiple scattering techniques. The results indicate that the energy distribution is continuous with a peak in the 35-40 MeV. region. The average energy is 40 MeV. It is concluded that a μ -meson decays to an electron and two neutrinos. These results are in reasonable agreement with those of other workers.

INTRODUCTION.

IN a recent publication from this laboratory (Miss R. H. Brown *et al.* 1949) in which applications of the new Kodak NT4 electron sensitive emulsions to a study of the cosmic radiation were described, it was reported that tracks of decay particles at minimum ionization had been observed to be emitted from mesons brought to rest in the emulsion. About 3 per cent of the mesons which stopped in the emulsion gave rise to a charged decay particle with range greater than 1000 μ . It was found possible to estimate the energy of these particles from observations on the deviations in the tracks due to multiple Coulomb scattering, and some preliminary results were given. These measurements have been continued, and we present here the results for 81 tracks with ranges in the emulsion greater than 1000 μ .

The tracks were observed to originate at the end of the range both of μ -mesons formed by the decay of π -mesons stopping in the emulsion, and from particles which, in less sensitive emulsions, we should have described as ρ -mesons. It has been shown that these mesons are identical with those which are commonly observed by means of cloud chambers and counters (Goldschmidt-Clermont *et al.* 1948; Camerini *et al.* 1948), and the present results can therefore be compared with those obtained by other methods..

METHOD.

For fast charged particles traversing a scattering medium, the mean angle of multiple scattering $\bar{\alpha}$ as defined by Williams (Williams 1939, 1940), can be expressed as

$$\bar{\alpha} = \frac{A}{m\beta^2c^2/\sqrt{(1-\beta^2)}},$$

where A is a constant dependent on the composition and thickness of

* Communicated by the Authors.

the medium traversed by the particle, m is the rest mass of the scattered particle, and β the ratio of the velocity of the particle to that of light, c . In the above equation, A may be calculated and $\bar{\alpha}$ is a measured quantity, leaving the mass and velocity of the particle unknown. Recent experiments (Hincks and Pontecorvo 1949, Leighton *et al.* 1949) strongly suggest that the mass of the charged decay particle is equal to that of an electron; further, all the published measurements on the energies of the decay particles from μ -mesons give values greater than 5 MeV. β must therefore be of the order of unity, and we may write:

$$\frac{m\beta^2c^2}{\sqrt{(1-\beta^2)}} \sim pc \sim E,$$

where p and E represent the momentum and total energy of the particle respectively.

Even if the mass of the scattered particle is considerably greater than that of an electron, the quantity $m\beta^2c^2/\sqrt{(1-\beta^2)}$ gives a measure of the total energy of the decay particle within the limits of error of our experiment. Thus for a particle of $10 m_e$ with a energy of 55 MeV.,

$$\frac{m\beta^2c^2}{\sqrt{(1-\beta^2)}} = 54.5 \text{ MeV.}; \quad p = 59.6 \frac{\text{MeV.}}{c};$$

and for an energy of 25 MeV.,

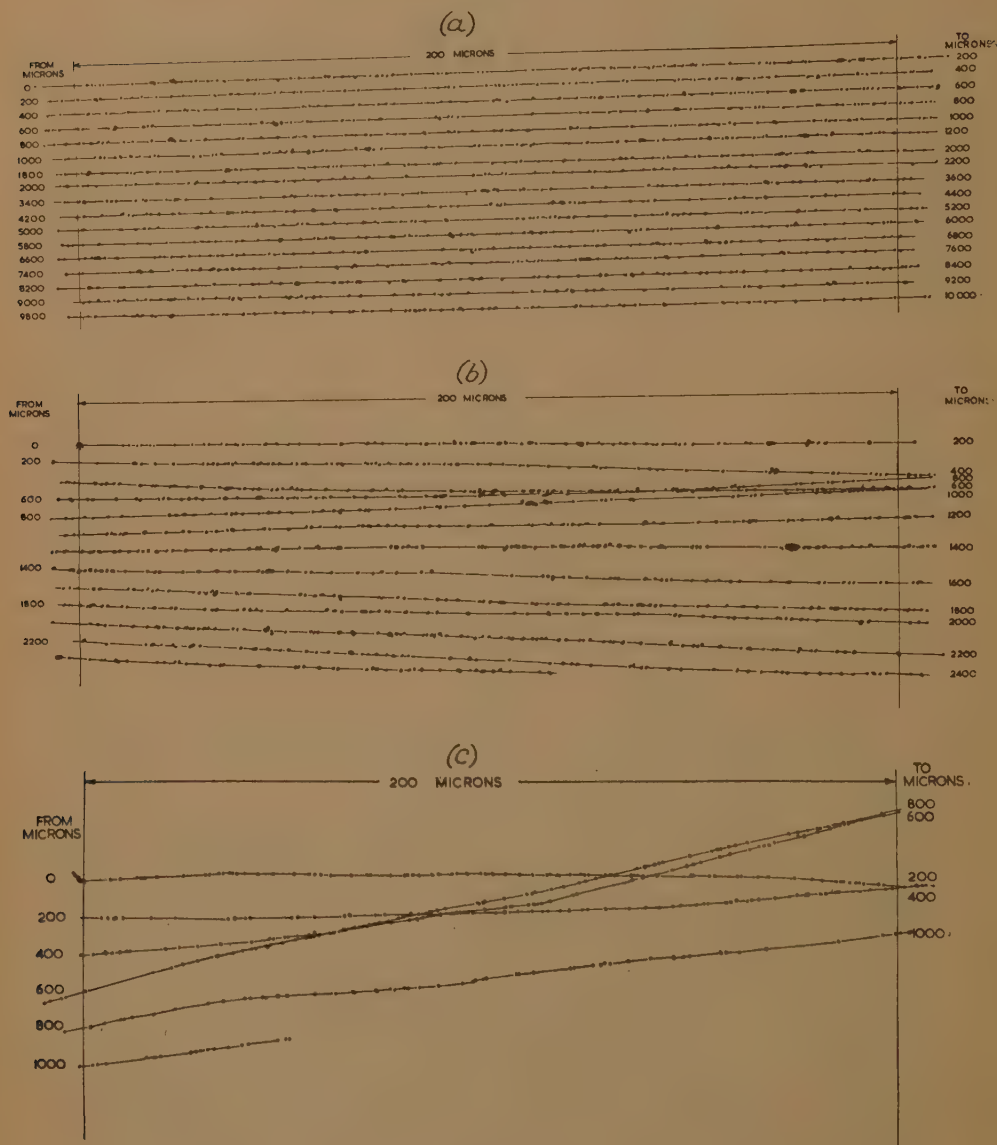
$$\frac{m\beta^2c^2}{\sqrt{(1-\beta^2)}} = 24 \text{ MeV.}; \quad p = 29.5 \frac{\text{MeV.}}{c}.$$

No serious errors in the values of the energy of the decay particles, deduced from our observations, can arise, therefore, from the assumption that their rest-mass is small.

Apart from a few improvements in technique, the experimental method employed for measuring $\bar{\alpha}$ was similar to that described in a previous communication (Goldschmidt-Clermont *et al.* 1948), to which reference should be made for details and nomenclature. The complete track was plotted on a single sheet of paper by means of an image thrown on a screen by a projection microscope. Typical examples of the drawings obtained in this way are shown in fig. 1. The drawing of each track was divided into segments, corresponding to an original length in the emulsion of 50μ , and the angular deviations, α , between the directions in alternate segments were measured. This procedure corresponds to employing a "cell" size of 100μ in the nomenclature of the previous paper.

The expected distribution of the values of α for the track of a particular particle is similar to that of a Gaussian curve to which has been added a "tail" at the higher angles, due to single scattering, which is governed by an inverse-cube law (Williams 1939, 1940). Since the number of readings which can be made along a particular track is not very great, a single large-scale deviation can have a serious influence on the value of $\bar{\alpha}$, the quantity on which the determination of the energy depends.

Fig. 1.



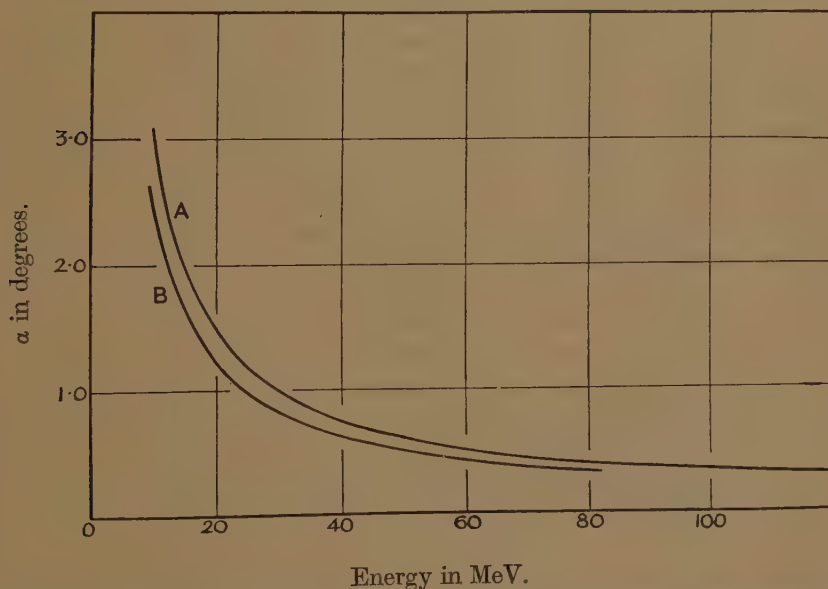
Facsimile tracings of three particles at minimum ionization. The secondary particle of a shower is given in (a), plots of 200μ length have been made at 800μ intervals illustrating the negligible distortion which can be achieved in the emulsion. Plots (b) and (c) show decay electrons of 40 MeV. and 9 MeV. respectively; the latter is from the μ -meson of a π - μ decay.

It is therefore necessary to adopt some form of "cut-off" procedure, excluding from the measurements all values of α above a certain minimum.

value. It has been shown (Goldschmidt-Clermont *et al.* 1948) that this "cut-off" angle is best chosen to be approximately equal to $4\bar{\alpha}_{CO}$, where $\bar{\alpha}_{CO}$ is the arithmetic mean of all values of α less than the "cut-off."

The preliminary measurements indicated that there is a considerable variation in the energy of the decay particles. Since the "cut-off" angle, as defined above, is approximately twice as great for the track of a 25 MeV. electron as for one of 50 MeV., a method was devised to enable the "cut-off" angle to be chosen according to the particular features of each track. This was done by determining the median of the measured values of α for each track. The value so obtained gives a measure of the energy of the corresponding particle which is less sensitive to the presence of a large-angle deviation in the track. With the approximate value of the energy so defined, the corresponding value of $\bar{\alpha}_{CO}$

Fig. 2.



Curve (A) represents the arithmetic mean value of all angles below the cut-off point for 100μ intervals under the conditions of our experiment, plotted against their energy. Curve (B) gives the corresponding values for the median as a function of energy.

could be deduced from a computed curve. The arithmetic mean of all values of α below the cut-off could then be determined, and hence a more precise value of the energy. Graphs showing the values of $\bar{\alpha}_{CO}$ and α (median) as a function of the energy of the particles are shown in fig. 2. They have been determined for "cells" 100μ long (50μ segments) and correspond to the particular conditions of our experiments, account being taken of effects due to "spurious scattering" and other minor corrections.

It will be seen from fig. 2 that for particles with an energy of ~ 50 MeV. $\bar{\alpha}$ is less than 1° . Any distortion in the emulsion would therefore lead to large errors in the estimated values of the energy. In "electron-sensitive" emulsions exposed to the cosmic radiation, however, large numbers of tracks are commonly observed due to particles moving at relativistic velocities which have traversed the emulsion. In the absence of distortion many of these tracks show no observable deviations. An examination of tracks of this type, which occur by chance near that of the decay-electron and which are approximately parallel to it, provides a good test for distortion. If some of these tracks are rectilinear, distortion may be assumed to be absent.

A second test to show that the present observations are in fact free from distortion and other causes was made by measuring the deviations in the tracks of high energy "shower" particles observed in the same plates, using methods similar to that employed for electrons. The results for a typical example, of length $10,000\ \mu$ (see fig. 1), are given in Table I., which shows the angular deviations for intervals of length $800\ \mu$. The results show that there is no curvature in the tracks due to distortion of the emulsion, or to imperfections in the optical system of the projection microscope.

TABLE I.

Interval	Mean Angle θ°	Deviation α°
0-200 μ	10.0	—
800-1000 μ	10.05	0.05
1600-1800 μ	10.025	0.025
2400-2600 μ	10.075	0.05
3200-3400 μ	10.0	0.075
4000-4200 μ	10.075	0.075
4800-5000 μ	10.1	0.025
5600-5800 μ	10.0	0.1
6400-6600 μ	10.075	0.075
7200-7400 μ	10.1	0.025
8000-8200 μ	10.2	0.1
8800-9000 μ	10.15	0.05

In Table I. θ represents the direction of a mean line passing through the grains in the corresponding segment of the track. The initial value of θ is arbitrary, and the deviation, α , represents the difference between consecutive values of θ .

The experimental errors in the determination of the energy, apart from those due to statistical fluctuations, are due to "spurious scattering," ~ 2 per cent, and the approximate method of analysis adopted, ~ 2 per cent, giving a total sum of ~ 3 per cent. The statistical error associated with each determination was equal to $56/\sqrt{n}$ per cent, where n is the number of statistically independent readings. The range of the decay electron is greater than $1000\ \mu$ in only about 10 per cent of the observed examples of meson decay, in emulsions of thickness $200\ \mu$. With tracks of length $1000\ \mu$, the probable error in the estimate of the energy, due to statistical fluctuations, is ~ 17 per cent. Although this

value is large, we have accepted all tracks for measurement of length greater than $1000\ \mu$, in order to secure a sufficient number for measurement.

There is a second reason for choosing $1000\ \mu$ as the minimum range accepted for measurement. If a larger value is chosen, there is a greater probability that electrons of low energy will be "lost" from measurement as a result of being so scattered that they leave the emulsion. This point is discussed in greater detail below.

For a given length of track, the smaller the chosen "cell" length, the smaller the errors in the determination of the energy, due to statistical fluctuations. There is a limit, however, to the reduction in "cell" size which can be permitted, for the true "deviations" in the track due to Coulomb scattering must be appreciably greater than those due to "spurious" scattering which arise from errors of measurement. The deviations due to Coulomb scattering decrease, and those due to errors of measurement increase as the "cell" size is reduced. The cell size we have adopted, viz. $100\ \mu$, represents a compromise based on these conflicting considerations and has been adopted in making a first estimate of the energy of all the decay electrons. In the case of the particles of lower energy a second determination can be made with advantage, using cells $50\ \mu$ long for electrons of energy less than 20 MeV.; and $70\ \mu$ long for particles of energy between 20 and 30 MeV. The values of the total probable errors (including experimental) associated with the final determinations of the energy, vary between 5 and 17 per cent, with a mean value of 11 per cent.

RESULTS.

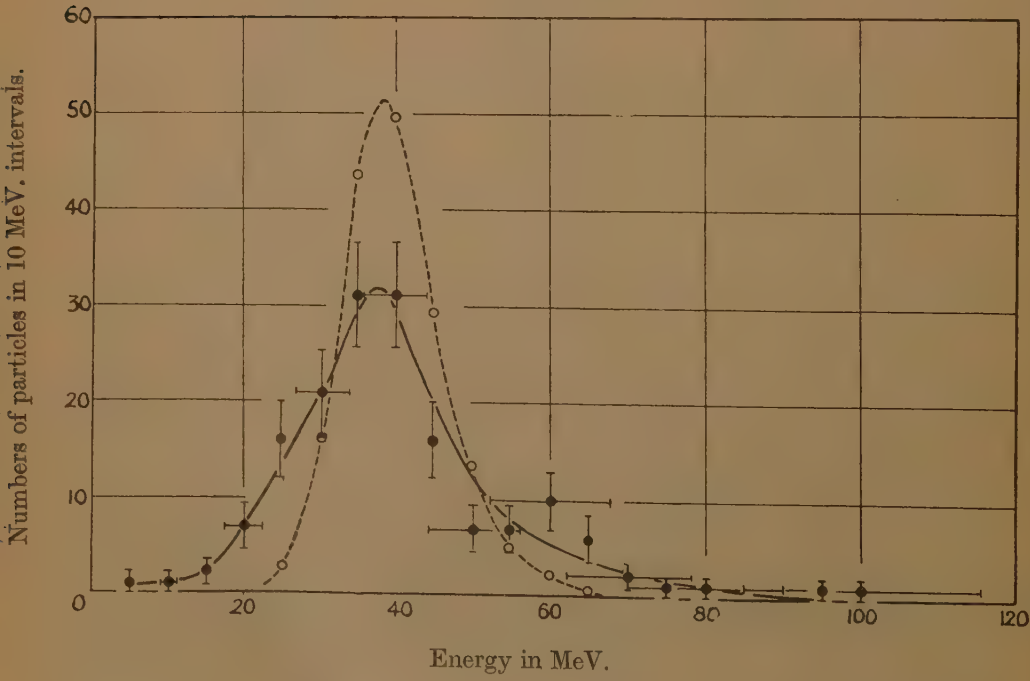
Table II. gives the values of the energy of 81 decay electrons together with the associated "probable" errors. The results are shown graphically in fig. 3. The solid line in this figure is drawn through points representing the number of particles of which the energy lies in successive intervals of 10 MeV., values being calculated for overlapping intervals so that the results are not statistically independent. The horizontal bars through the experimental points indicate the mean value of the probable errors in the determination of the energy of the individual particles in the corresponding interval of energy; and the vertical bars the statistical errors in the observed frequency of occurrence of particles with an energy in the corresponding intervals. The average value of the energy of the decay electrons is 40 MeV., and the peak in the observed distribution lies between 35 and 40 MeV.

If the energy of the decay electron were constant, and equal to 40 MeV., we should expect to observe, in the conditions of our experiments, a distribution similar to that indicated by the broken line in fig. 4 (Deming and Birge 1934). The points in this curve represent the expected numbers of tracks for the different energy intervals, the width of the peak being due to the statistical fluctuation in the values of $\bar{\alpha}_{CO}$. The disparity between the two curves shows that the energy of the decay

TABLE II.
Energy of decay electrons in MeV.

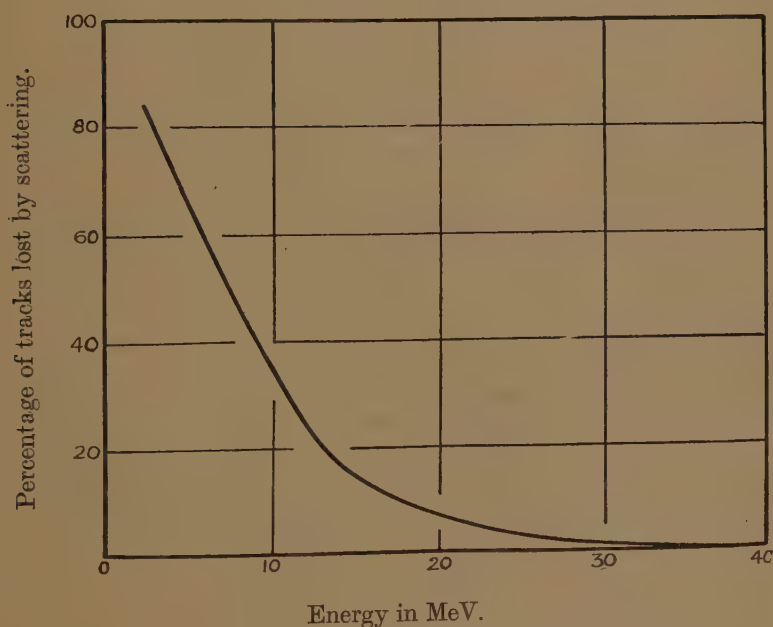
* 9.0±1.0	32.0±4.5	38.0±2.0	48.0±6.0
15.0±2.0	32.5±4.0	38.0±2.0	49.0±6.0
16.5±2.0	32.5±4.5	38.0±6.0	51.0±8.0
21.0±2.5	33.0±4.0	38.5±2.0	51.0±5.5
21.5±2.0	33.0±4.5	39.0±5.0	56.0±9.0
23.0±2.0	33.0±5.0	39.0±4.0	57.0±7.0
24.0±3.0	34.0±4.0	39.0±6.0	58.0±9.0
24.0±3.0	34.0±5.0	39.0±6.0	59.0±10.0
25.0±3.0	35.5±5.5	40.0±4.5	59.0±4.0
25.0±3.0	36.0±3.5	40.0±2.5	59.5±8.0
25.0±1.5	36.0±3.5	*40.5±5.0	60.0±8.5
25.0±3.0	36.0±4.0	42.0±4.0	60.0±6.5
26.5±3.5	36.0±4.5	42.5±5.0	63.0±11.0
28.0±2.0	36.5±4.0	42.5±4.5	64.0±6.5
*28.5±4.0	37.0±3.0	43.0±5.5	66.0±7.0
28.5±1.5	37.0±4.5	43.0±4.0	67.0±9.0
29.0±4.0	37.0±5.5	*43.0±5.5	75.0±10.0
29.0±3.0	37.0±5.5	44.0±6.0	99.0±13.0
29.0±4.0	37.5±2.5	45.5±4.5	
30.5±4.0	39.0±5.0	47.0±6.0	
32.0±4.0	38.0±4.0	48.0±6.0	

Fig. 3.



electrons is not unique; that the values are distributed in an energy spectrum. The strongest evidence for this conclusion is the observation of several particles of energy less than 20 MeV., for which it would be very difficult to account in terms of statistical fluctuations alone. The results are in agreement with those of other workers using other methods (Leighton *et al.* 1949; Steinberger 1949). It may be noted that in those cases in which a μ -meson can be definitely identified, since it is observed to originate by the decay, in the emulsion, of a π -particle, wide differences in the values of the energy of the decay electrons are also found. The values of the energy of these particular particles are marked by an asterisk in Table II.

Fig. 4.



We have mentioned above that there is a certain tendency to underestimate the number of decay electrons of low-energy, owing to the fact that particles which would remain in the emulsion for a distance of 1000μ if their trajectories were strictly rectilinear, may be so deviated as a result of Coulomb scattering that they escape prematurely. This tendency is more pronounced for particles of low energy which are more heavily scattered. It is not completely compensated by the opposite tendency, in which a particle remains in the emulsion due to scattering, when, if the trajectory had been strictly rectilinear, it would have "escaped". A graph showing the variation of the percentage "loss" of tracks, as a function of the energy of the particles, is shown in fig. 4.

The curve has been calculated for an emulsion of thickness 200μ for the particular case in which only tracks longer than 1000μ are accepted for measurement. No account of this effect has been taken in determining the experimental distribution given in fig. 3. By making an appropriate correction to these results the weight of the evidence indicating a "spread" in the values of the energy of the decay electrons would be increased.

DISCUSSION.

The mean energy of the electrons produced by the decay of μ -mesons, as deduced from our experimental results is 40 MeV. It will be seen from the broken line in fig. 3 that a group of electrons homogeneous in energy, would appear in our experiments as a skew distribution, with a "tail" at higher energies, and a much sharper "cut-off" on the side of lower energies. Because of this particular feature of our method we believe that the true value of the mean energy is a little below 40 MeV. Further, our results give no indication that the observed distribution is due to a superposition of the effects due to the presence of several homogeneous groups of electrons of different energy. The results therefore suggest that in the decay of a μ -meson the electron is accompanied by the emission of two neutral particles of small rest mass; for it is reasonable to assume that in such a three-body disintegration the mean energy of any one of the three particles will be equal to one third of the available energy, viz. to 36 MeV., taking the rest-mass of the μ -meson to be $216m_e$.

The work of Hincks and Pontecorvo (1948) has shown that high energy photons are not emitted during the decay of the μ -meson. Further, the majority of the experiments made by other workers, and our own results, indicate that the maximum energy of the emission of the electrons is ~ 55 MeV. This result gives support for the view that the neutral particles, emitted together with the electron, are of small rest-mass. This makes it probable that they are neutrinos, so that we may represent the process of decay:

$$\mu \rightarrow e + \nu_0 + \nu_0.$$

In support of this conclusion both Marshak (1949) and Serber (1949) have pointed out that if the π -meson decays according to the scheme $\pi \rightarrow \mu + \nu_0$, as now appears to be established, then it follows that a μ -meson should decay into an electron and two neutrinos. It is therefore probable that the μ -meson has spin $\frac{1}{2}$ and obeys Fermi statistics, whilst the π -meson has spin 1 or 0 and obeys Bose statistics.

Assuming a mass of $216m_e$ for the μ -meson, the maximum possible energy of the decay electron is 55 MeV. In fig. 3 it will be seen that there are several values well above this limit. It seemed to be important to establish whether these electrons of apparently greater energy were due to statistical fluctuations in the deviations in the tracks; or,

alternatively, whether they were the product of the decay of mesons more massive than the μ -particles. We therefore made measurements on the tracks of the parent particles in order to determine their mass by the scattering method (Goldschmidt-Clermont *et al.* 1948). Because of the large statistical errors inherent in the method when it is applied to the track of a single particle, we deduced a mean value for the mass of all particles which produced an electron of which the apparent energy was greater than 60 MeV. The value thus obtained was $204 \pm 19 m_e$. Our evidence is therefore consistent with the assumption that all the electron tracks we have examined were produced by the decay of μ -mesons.

In view of the above result we attribute the electrons with an apparent energy about 55 MeV. to statistical fluctuations. They are sufficiently numerous to suggest that a considerable fraction of the decay electrons have an energy in the interval between 45 and 55 MeV.; that the true distribution curve shows a sharp "cut-off" at 55 MeV. and does not approach this limit asymptotically. This conclusion is in accord with certain theoretical distributions deduced by Tiomno, Wheeler and Rau (1949); see also Michel (1949). These authors have made an exhaustive theoretical analysis of the μ -decay, and have found that for many types of coupling between the different particles, and in particular the vector and pseudo-scalar varieties, there is a finite probability of producing electrons with the maximum possible value of the kinetic energy.

All our measurements made hitherto have been done with emulsions 200μ thick. It is now possible to process 600μ emulsions and it seems likely that in the near future it will be possible to work with emulsions of considerably greater thickness. These will allow us to make measurements on tracks of much greater length in sufficient numbers and thus to increase greatly the statistical weight of the measurements. It therefore seems probable that we shall be able to determine, with much greater precision than has thus been achieved hitherto, the form of the energy-spectrum of the electrons produced by the decay of μ -mesons.

ACKNOWLEDGMENTS.

We have pleasure in thanking Professor C. F. Powell, F.R.S., for his continued interest and encouragement during the course of this work: U. Camerini and Dr. D. M. Ritson for helpful discussions, and Dr. H. Heitler for technical assistance.

REFERENCES.

- BROWN, Miss R. H., CAMERINI, U., FOWLER, P. H., MUIRHEAD, H., POWELL, C. F., and RITSON, D. M., 1949, *Nature*, **163**, 47.
 CAMERINI, U., MUIRHEAD, H., POWELL, C. F., and RITSON, D. M., 1948, *Nature*, **162**, 343.
 DEMING, W. E., and BIRGE, R. T., 1934, *Rev. Mod. Phys.*, **6**, 119.
 GOLDSCHMIDT-CLERMONT, Y., KING, D. T., MUIRHEAD, H., and RITSON, D. M., 1948, *Proc. Phys. Soc.*, **61**, 183.
 HINCKS, E. P., and PONTECORVO, B., 1949, *Phys. Rev.*, **75**, 698; 1948, *Phys. Rev.*, **75**, 257.

- LEIGHTON, R. B., ANDERSON, C. D., and SERIFF, A. J., 1949, *Phys. Rev.*, **75**, 1432.
 MARSHAK, R. E., 1949, *Phys. Rev.*, **75**, 700.
 MICHEL, L., 1949, *Nature*, **163**, 959.
 SERBER, R., Feb. 1949, *Phys. Rev.*, **75**, 1459.
 STEINBERGER, J., 1949, *Phys. Rev.*, **75**, 1136.
 TJOMNO, J., WHEELER, J. A., and RAU, R. R., 1949, *Rev. Mod. Phys.*, **21**, 144.
 WILLIAMS, E. J., 1939, *Proc. Roy. Soc. A*, **169**, 531; 1940, *Phys. Rev.*, **58**, 292.

CXVI. *The Interpretation of X-Ray Absorption Spectra of Solids.*

By Y. CAUCHOIS * and N. F. MOTT †.

[Received September 27, 1949.]

[Plates XXIV. & XXV.]

ABSTRACT.

A discussion is given of some recent experimental results on the fine structure of the X-ray absorption edges of solids. Recent work has shown the existence for many solids of a pronounced maximum (raie blanche) at the absorption edge. For K-absorption this may be due to one of two causes: (i) a high density of normally unoccupied states, with p symmetry in the neighbourhood of the absorbing atom; this is for example the case for nickel in oxides; and (ii) the formation of exciton levels. Exciton levels are defined as energy levels which are not formed until the electron is removed from the X-ray level, and are caused by the field round the point where the charge is missing. They are to be expected in insulators but not in metals. The experimental evidence is discussed in the light of this theoretical model, with special reference to the semi-metals gallium and arsenic.

§ 1.

THE absorption spectrum of a free atom, for X-rays as for visible light, consists of a series of lines leading up to a series limit; in the case of X-rays the lines are considerably broadened on account of the short life-time, due to Auger effect, of the ionized state of an X-ray level (Richtmyer, Barnes and Ramberg 1934). An example is the K absorption of argon measured by Parratt (1939). It seems fairly certain from the theoretical point of view that the X-ray absorption spectra of solids which do not show metallic conduction ought to consist of a similar series of lines leading to a limit. This is for the following reason: the series limit corresponds to a transition in which the electron from the X-ray (*e.g.* the K) level is ejected directly into the conduction band (transition (*a*) of fig. 1). But as soon as the electron has left the K level, it leaves a residual positive charge on the K shell; and owing to this there are formed, below the bottom of the conduction band, the series of so-called

* Lab. de Chimie physique, Université de Paris.

† H. H. Wills Physical Laboratory, University of Bristol ‡.

‡ Communicated by Prof. N. F. Mott, F.R.S.

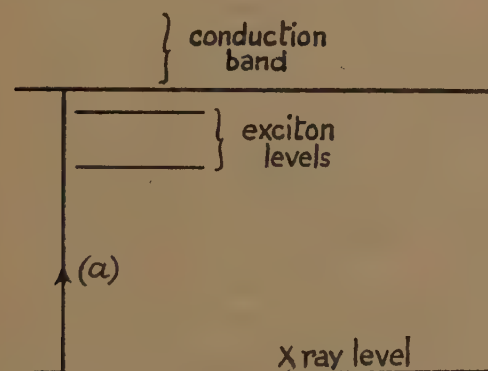
exciton levels into which the electron may make a transition. The existence of these levels has its origin in the fact that the field round the positive charge may, at large distances, be taken to be of the form $e/\kappa r^2$, so that the additional potential energy of an electron moving in it is

$$V(r) = -e^2/\kappa r.$$

κ is the dielectric constant at high frequencies (square of the refractive index). The Schrödinger equation for an electron moving in this field, even when superimposed on the periodic field of the lattice, must have a series of quantized values of the energy leading to a series limit.

In general these levels should be much closer together than for the isolated atoms. Consider, for example, the case of solid argon, where, after absorption of X-rays by a K-electron, the electron will be in a $4p$ orbit. The $4p$ orbit has a much greater radius than that of the *neutral* atom, and thus than the interatomic distance in the solid. In the solid,

Fig. 1.



Energy levels of a non-metal.

therefore, the potential in which the electron moves, in addition to the periodic field of the lattice, is as already stated $-e^2/\kappa r$. In the formula for the interval between the energy levels, κ will enter in the denominator squared; one has only to remember the Bohr formula $2\pi^2me^4/\hbar^2$ to see this. Therefore one expects that in general these exciton levels will be separated from the series limit by an interval smaller by a factor κ^2 than for the isolated atom.

The ultra-violet absorption spectrum of the alkali-halides has been discussed by Mott and Gurney (1940) in terms of this model; the absorption is due to a transition of the outer p -electron of the halide ions into an exciton level, supposed to have s symmetry relative to the halide ion, or directly into the conduction band. The separation between the exciton level and the series limit is of the order of 2 electron volts. For substances with higher dielectric constant one would expect it to be still smaller, perhaps a fraction of an electron volt. With separations of this order, for the K absorption of solid substances one would expect the exciton level to be difficult if not impossible to observe. On the other

hand, if there are vacant p -orbitals in the absorbing atom (e. g. in arsenic, solid halogens, etc.), a K -electron ejected into an exciton level may well remain within its own atom, so that the argument given above about the reduction in the force between the electron and the positive charge no longer applies. In these cases, therefore, the exciton level may well be several electron volts below the conduction band and therefore observable in X-ray absorption spectra.

It is the principal purpose of this paper to examine the experimental material on X-ray absorption and to see what evidence there is for the existence of these exciton levels. This will be done in § 2. We shall first consider some other details of the absorption spectrum of non-metals, and then contrast the expected behaviour of metals.

In a free atom, whether one is dealing with optical or X-ray absorption, the *theoretical* value of the absorption coefficient at the series limit is finite, not zero. On theoretical grounds, also, the same should be true for an insulator. This result, whether for the free atom or for the insulator, seems at first sight surprising, because $N(E)$, the density of states in a band or in a vacuum, tends to zero as $E^{\frac{1}{2}}$ with E , and the intensity of absorption $I(E)$ is proportional to $N(E)$. If $p(E)$ is the transition probability, we may write

$$I(E) = \text{const. } p(E)N(E),$$

and it thus follows that $p(E)$ must tend to infinity as $E^{-\frac{1}{2}}$ as $E \rightarrow 0$. That this is so is actually a consequence of the Coulomb field of potential $-e^2/\kappa r$ superimposed on the lattice field in which the electron moves when ejected into the conduction band. The transition probability $p(E)$ is proportional to the square of an integral of the type

$$\left| \int \psi_f^* \frac{\partial}{\partial x} \psi_K d\tau \right|^2,$$

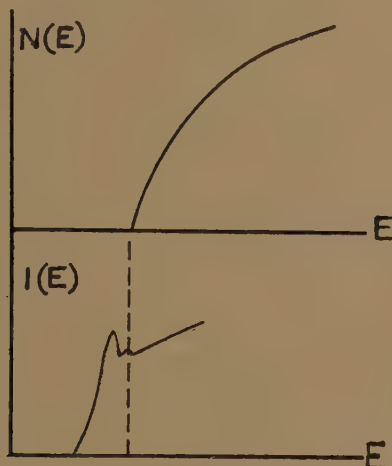
where ψ_K is the wave function in the initial K state and ψ_f in the final state. On account of the attractive Coulomb field, the value of $|\psi_f|^2$ at the origin (and thus in the region occupied by the K shell) tends to infinity as E tends to zero—or rather as the velocity of the electron tends to zero (cf. for example Mott and Massey 1949).

All these arguments ought to apply not only to the bottom of the first empty band but to the bottom of all the other bands as well. The velocity (group velocity $\partial E/\partial k$) is zero here, just as at the bottom of the first band. Thus one expects a finite absorption coefficient at the bottom of *any* band, and exciton levels below it. The spectrum should thus be as shown in fig. 2 (for reasons to be given below this applies only to non-metals). Probably experimental resolution would not be sufficiently great to show up the exciton levels, but what one expects is a *sharp* beginning to each absorption minimum in the Kronig structure, as indicated in fig. 3 (curve (a)). In figs. 4 and 5 we shall show at least two insulators (Millerite (NiS) and Ni_2O_3) where these predictions seem to be borne out by experiment.

It must be admitted that there are many reasons why this behaviour

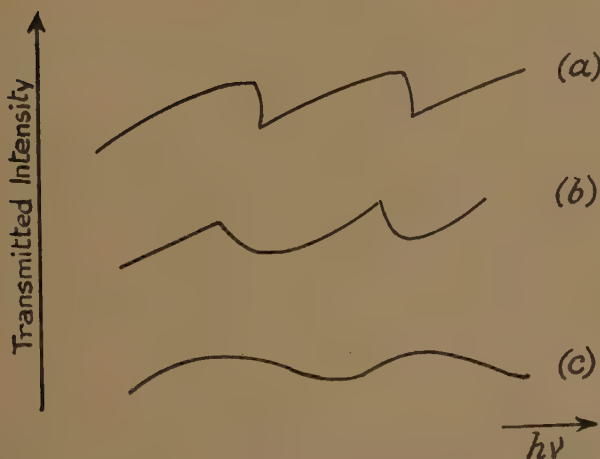
should be obscured, particularly the very short life-time of the excited states due to collisions between electrons in the conduction band and those in the lower filled bands. No quantitative theoretical estimates have been made of these life-times, and we have no idea how they vary from

Fig. 2.



Density of states and absorption coefficient at the beginning of a band in a non-metal.

Fig. 3.



- (a) Expected type of absorption for non-metal.
- (b) Metal.
- (c) Effect of finite breadth of lines.

one compound to another. Also the width of the X-ray levels due to Auger effect may be important. The examples of the Kronig fine structure of Cu_2O and CuO and ZnO shown in figs. 12, 13 and 14 illustrate how, for instance, the sharpness of minima in the absorption can vary from one substance to another.

The position in metals from the theoretical point of view is quite different. It is characteristic that a positive charge in a metal is screened by the surrounding electrons, so that the potential energy of an electron in its field is no longer to be represented by $-e^2/r$ but by

$$V(r) = -e^2/r \exp(-qr), \quad \dots \dots \dots (1)$$

where q^{-1} is of the order 1 to 2Å. This principle is used by Mott and Jones (1936) to calculate the additional resistance due to additions of Zn etc. to Cu, and other applications will be mentioned below. In considering the X-ray absorption edge of metals, then, the field in which the electron moves after the absorption process consists of the lattice field with in addition the field derived from the potential (1). Now for an electron moving in a field of the type (1), there will be at most a *finite* number of stationary states, and if q is large enough, no bound stationary states at all.

One of us (Mott 1949) has recently given reasons for thinking that, in general, in metals q is so large that no stationary states exist. The reason is that it is characteristic of a metal that the electrons are free, and thus that the electrons and positive ions do not combine to form pairs from which the electron is incapable of moving. This being the case, one would not expect exciton lines at the absorption edges of metals; also the transition probability $p(E)$ would not tend to infinity at the bottom of a band. Thus the absorption spectrum of a metal in the Kronig region may be expected to be much less marked at the beginning of each band, following more truly the $N(E)$ curve (cf. fig. 3 (curve (b))). Though here probably too, the short life-time of the excited states will make everything much less distinct, so that the curve in fact looks like fig. 3 curve (c).

Although one does not in general expect exciton lines for metals, it is probable that arsenic, antimony and bismuth may be exceptions. These (cf. Mott and Jones 1936) have a very small number of free electrons and positive holes, which are prevented from forming pairs only because of their very low effective mass (Mott 1949). This exceptional state of affairs will be broken down as soon as a vacancy is formed in the X-ray level. We return to the consideration of this point in § 3.

§ 2.

We shall now discuss the experimental material with a view to seeing what evidence there is for the existence of exciton levels. One of us in a recent note (Cauchois 1949 a), has emphasized the frequent occurrence of a "white line," (French raie blanche, denoted in what follows by R.B.), or maximum of the absorption coefficient, immediately on the high frequency side of the K and L absorption edges of mineral compounds, and its absence in most metals. In this section we shall discuss certain cases in which this white line can probably be interpreted as due to the formation of exciton levels in the field of the positive charge left by the ejected electron.

An R.B. *not* due to an exciton level can only arise if there are electronic states of the required symmetry with a very high density of levels at the absorption edge. The best known case is that of the transition metals, where, owing to the small radius and consequent small overlap of the *d*-orbitals, there is a high density of empty states of predominantly *d* symmetry. Consequently an R.B. appears in the L_2 and L_3 absorption edges of elements such as W (Veldkamp 1935), and in the L_3 of Pt (Cauchois and Manescu 1940, for an explanation of its absence in the L_2 edge, *cf.* Mott 1949).

A similar explanation, due originally to Coster and Kiestra (1948) can account for the R.B. obtained at the metal K edge of many metallic oxides. According to these authors the empty *p*-band (of symmetry *p* in the neighbourhood of the metal ion) becomes narrow and with wave functions less and less hybridized as more oxygen is added and the metal ions move further from each other. This is shown in figs. 12, 13 and 14, which illustrate results obtained in the Paris laboratory for ZnO, CuO and Cu_2O . Zinc oxide shows a very marked R.B., as does also CuO; the far more covalent Cu_2O , in which the copper ions are closer together, does not show it. Similar results have been obtained for the oxides of nickel (Cauchois and Manescu 1949, to appear in the *Journ. de Chimie Phys.*; see also Cauchois 1949 b).

Turning now to cases where the R.B. observed is probably due to an exciton, we have first the K-absorption of arsenic shown in fig. 8 obtained by Hulubei and Cauchois (1940). As already stated, it seems to us unlikely that there should be any particularly high density of *p*-states at the surface of the Fermi distribution in this metal; and probably exciton formation occurs. For the elements of similar structure, no R.B. has been observed for the K-absorption of Sb, perhaps on account of insufficient resolution, but Sandström (1935) found one for the L_1 edge; bismuth shows one for L_1 (Cauchois 1942). These metals are discussed further in § 3.

Gallium shows a very marked R.B. in both the solid and liquid states (Hulubei and Cauchois 1940, M. Vidal 1949) (see also fig. 6). Without attempting a detailed explanation, for the behaviour of the liquid, we imagine that the situation is somewhat similar to that for arsenic (*cf.* § 3).

As expected, there is no R.B. in the K-absorption of metals such as Cr, Mn, Fe, Co, Ni, Cu (*cf.* fig. 9 for Ni); the case of zinc is doubtful; some minima have been observed by one of us (Y.C.) which seem to disappear in zinc of the highest purity. In tin no R.B. has been observed in the K-absorption, though Sandström finds one for L_1 ; the difference may be due to insufficient resolving power for the K-absorption.

A very interesting example of an R.B. at an absorption edge which is almost certainly an exciton level is found in the K-absorption of arsenic in NiAs (fig. 10). Here, unlike the case for NiO, there is no R.B. at the metal edge (fig. 11), or, at any rate, no minimum of importance in comparison with the usual fluctuations of intensity. Without giving a detailed explanation, it appears likely enough that the highly covalent nature of this compound results in heavy coupling between the *vacant p*-orbitals

of Ni, thus preventing a high density of vacant p states and so an R.B. at the Ni K-edge. In any case the absence of an R.B. for Ni shows that there is no dense band present *until* the arsenic K-level has lost its electron; in other words, we have to do here with an exciton level.

It is interesting that if one compares the R.B. for As in NiAs with that for As in the elementary form, one finds that the former is probably very slightly displaced towards low frequencies. For a compound one usually obtains a displacement towards large frequencies. This perhaps suggests that we have to do with excitons in NiAs at any rate. The nickel discontinuity on the other hand is displaced, relative to that of the metal, in the other direction*.

A recent observation by one of us (Y.C.) of the rubidium K-edge of RbCl is illustrated in fig. 7. This shows a very sharp R.B. of the type expected. Finally, we show in fig. 4 the fine structure of the absorption of natural NiS (Millerite). This non-metal shows, in contrast to Ni, an absorption very much of the type anticipated for insulators (fig. 3) with sharp increases in the absorption coefficient and relatively gentle decreases. It does not seem true, however, that this is a general feature of non-metals.

§ 3. THE ABSORPTION SPECTRA OF METALS SUCH AS ARSENIC AND GALLIUM.

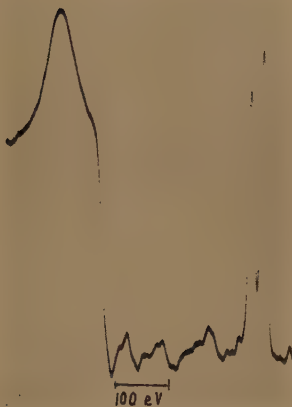
These are of particular interest. There is no reason to believe that the strong R.B. observed could be due to any high density of p states existing before the K level was ionized. On the other hand it seems impossible that an exciton level of the type described above could be formed, because there is no zone of forbidden energies where it could be placed.

We believe that the proper description of the process is somewhat as follows: the metals discussed are those in which the overlap between the full band and the empty band is very small; they are only just metals. Consequently the wave-length of the conduction electrons is very large, many multiples of the lattice constant. Consider then the wave-function of the p -electron that has been ejected from the K-level of an arsenic atom. The wave-function, we may suppose, is not too different from that of atomic arsenic in the same state of ionization. But in the metal the energy of the electron must be considered as at the surface of the Fermi distribution; therefore the wave-function must join up, at the boundary of the atomic polygon surrounding the arsenic atom considered, with the wave-functions of the conduction electrons. These, as already seen, have very long wave-length. A phenomenon therefore occurs familiar in the theory of scattering of slow neutrons; a wave of long wave-length has to be fitted to a "potential hole" in which there is a virtual quantized energy level at zero energy. Under these conditions, as is well known, a kind of resonance occurs and the amplitude of the wave-function within the "hole" becomes large compared with that of the

* One takes here for the metal the *first* of the two discontinuities, K_1 of fig. 9, in agreement with the usual interpretation.

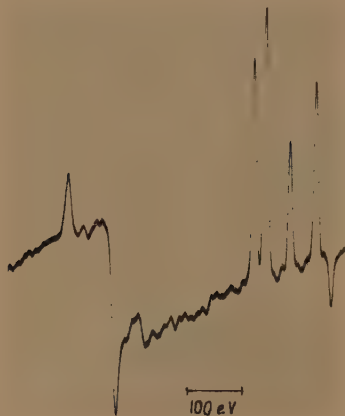
FIG. 4.

FIG. 5.



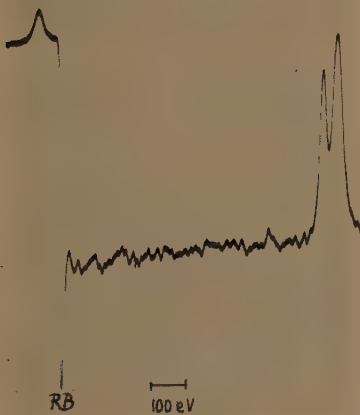
Nickel in NiS (millerite).

FIG. 6.



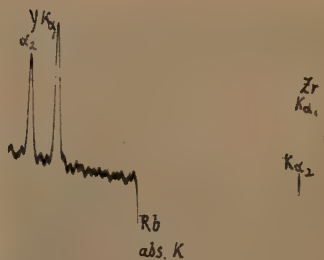
Nickel in Ni₂O₃.

FIG. 7.



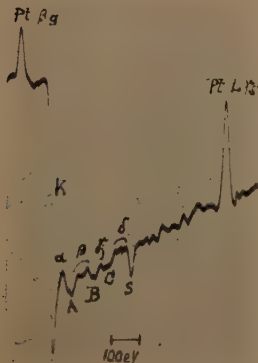
Metallic gallium (solid) at a temperature near melting point.

FIG. 8.

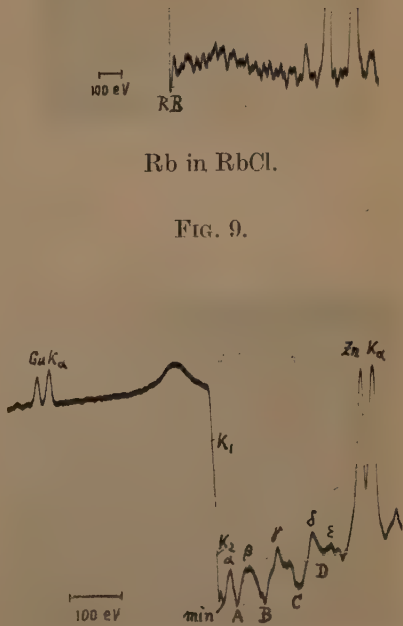


Rb in RbCl.

FIG. 9.

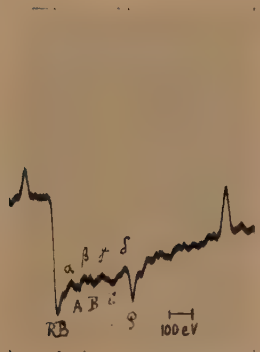


Metallic arsenic.



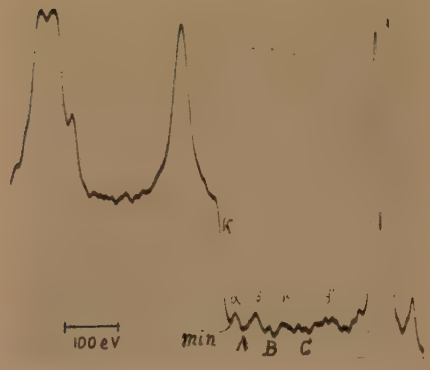
Metallic nickel.

FIG. 10.



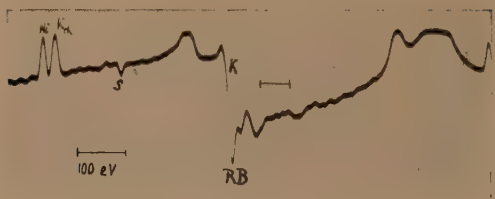
As in NiAs.

FIG. 11.



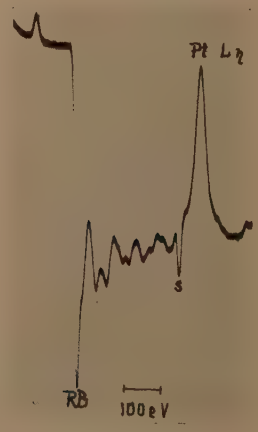
Nickel in NiAs.

FIG. 12.



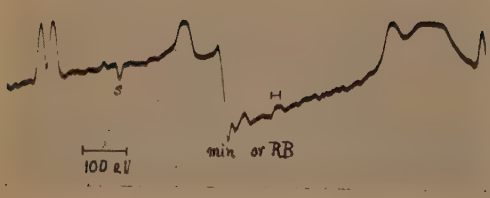
Copper in CuO.

FIG. 14.



Zinc in ZnO.

FIG. 13.



Copper in Cu₂O.

waves of long wave-length outside. Details of this will be found in Mott and Massey 1949. This, then, is the mathematical description of the cause of the R.B., since the transition probability $p(E)$ is determined by the amplitude of the wave-function within the "hole." Physically therefore, we may say that when a substance is only just metallic, it is possible for it to behave as though it were a non-metal and give an exciton type of absorption characteristic rather of the free atom than of the solid.

It is hoped to publish a more detailed mathematical investigation of this phenomenon in the near future.

§ 4. DETAILS OF EXPERIMENTAL MATERIAL.

The absorption spectra which have been used to illustrate the theoretical part of this paper were obtained under the following conditions : the small demountable X-ray tube in the laboratory of Physical Chemistry of the University of Paris can be used up to 50 kV. Most of the absorption spectra were, however, obtained at a much lower voltage, less than twice the excitation potential ; the reason for this is to avoid exciting radiation with a frequency two or three times that of the edge under consideration, which could be superimposed on the actual spectrum because of reflections of a higher order. This precaution is not absolutely necessary when the second order reflection is very weak (as in the case of the (100) plane of mica).

The spectrograph consists of a curved sheet of mica, and use is made of the ($h0l$) reflections through the sheet. The radius of curvature is 40 cm. The 100 reflection give the best contrast for the Kronig structures ; but the first, second and third order reflections on (201) allow better resolution and make it possible to effect more exact measurements, especially of the actual edges, with the help of narrow K lines of well known wave-lengths.

The absorbing screen was generally placed between the crystal and the photographic plate in such a way as only to receive the scattered radiation. In these conditions the photochemical action of the X-rays during the exposure remains very weak. If, on the other hand, one places the screen between the window of the tube and the crystal in the intense incident ray, the chemical change may be important. Further, the fluorescent spectrum is superimposed on the absorption spectrum and could be a nuisance.

The materials for which the absorption was measured were for nickel rolled sheet 0.1 mm. in thickness ; for other substances, a homogeneous layer of powder spread out on greasy cigarette paper or Scotch tape. The mass per unit area of each element of which the absorption was obtained was, as far as possible, the same for each case of different chemical composition. The photographs were taken on Ilford or Gevaert plates and examined with microphotometers. The exposure times, reduced to a 5 milliamperes run, varied from 20 minutes to some hours.

Details of Figures 4 to 13 and of the Table.

All the microphotometer curves show the K absorption of the element indicated, the frequency increasing from left to right. We have shown in the accompanying table the wave-lengths of some typical absorption curves (Ni, As and Rb). We have added, too, some measurements made on metallic arsenic in 1940 with a resolving power greater than that used in 1949. These measurements give a more complete description of the structure up to 50 eV. from the absorption edge, but the structure for higher energies was not investigated. The strong absorption of arsenic in NiAs made the use of a higher resolving power difficult.

The white lines (maxima in the absorption) marked with S are due to multiple reflections in the analysing crystal; they thus give a measure of its resolving power.

All the curves reproduced here are obtained with the (100) reflection of a curved mica sheet of radius 40 cm.

In the table, maxima are printed in bold-faced type.

Nickel metal (fig. 9).

	X.U.	eV.
K ₁	1485.0 ₉	0
K ₂	1483.1 ₇	10.8
1st min.	1482.0 ₅	17.6
α	1478.9	35.2
A	1476.0	51
β	1474.6	60
	1473.8	64
	1472.5	72
B	1467.8	97
γ	1463.1	125
	1461.6	134
	1459.0	149
C	1455.8	168
δ	1451.7	191
D	1448.0	213
ε	1445.0	230
E	1441.3	252
η	1439.0	267

Nickel in NiAs (fig. 11).

	X.U.	eV.
K	1484.0 ₂	0
1st min.	1481.9	12.2
α	1479.7	24.4
A	1476.7	42
β	1473.0	62
B	1468.5	88
γ	1464.2	114
C	1456.4	158
δ	1451.3	188

Arsenic in NiAs (fig. 10).

	X.U.	eV.
K edge	1043.0 ₄	0
Bottom of R.B.	1041.8	13.5
Max.	1040.2	31.2
Min.	1039.2	43.3
α	1037.5	62
A	1035.3	88
β	1032.9	115
	1031.0	138
B	1029.3	157
γ	1026.9	186
C	1023.4	226
δ	1021.2	252
	to 1019.3	275

Arsenic metal (fig. 8).

	X.U.	eV.
K edge	1042.8 ₆	0
Bottom of R.B.	1041.2	19.
α	1038.8	46
A	1035.5	84
β	1033.7	106
	1030.9	138
B	1028.1	171
γ	1026.3	191
C	1024.5	213
δ	1022.6	234
	to 1018.0	290

Measurements by Hulubei and Cauchois in 1940, with a greater resolving power than that used here, showed a fine structure of the R.B. of arsenic metal, as shown below.

Arsenic metal	X.U.	eV.
K edge	1042.8 ₆	0
R.B.	1042.4 ₁	5.0
dark line	1041.8 ₇	11.2
minimum	1041.2	18.8
dark line	1040.6	25.5
minimum	1039.7	36

NOTE.—The difference between the edges of Ni in NiAs and nickel metal is +6.0 eV.; between those of As in NiAs and arsenic metal is -2.0 eV.

Rubidium in RbCl (fig. 7).

	X.U.	eV.
K edge	813.9 ₄	0
minimum	813.6	5.4
maximum	813.1	14.9

ACKNOWLEDGMENTS.

We would like to express our thanks to Mlle. I. Manescu for her help in measuring the plates and to Professors Hocart and Goldsztaub and M. Weil of the University of Strasbourg for providing us with samples of NiAs and other materials for this investigation.

REFERENCES.

- CAUCHOIS, Y., and MANESCU, I., 1940, *C.R. Acad. Sci., Paris*, **210**, 172.
 CAUCHOIS, Y., 1942, *Cahiers de Physique*, **8**, 25; 1949 a, *C.R. Acad. Sci., Paris*, **228**, 1720; 1949 b, *J. de Chimie physique*, **46**, 307.
 COSTER, D., and KIESTRA, D., 1948, *Physica*, **14**, 175.
 HULUBEL, H., and CAUCHOIS, Y., 1940, *C.R. Acad. Sci., Paris*, **211**, 316; *Disq. Mat. Phys.* **I**, **3-4**, 467.
 MOTT, N. F., 1949, *Proc. Phys. Soc.*, **62**, 416.
 MOTT, N. F., and GURNEY, R. W., 1940, *Electronic Processes in Ionic Crystals*.
 MOTT, N. F., and JONES, H., 1936, *Theory of the Properties of Metals and Alloys*.
 MOTT, N. F., and MASSEY, H. S. W., 1949, *The Theory of Atomic Collisions*, 2nd ed., p. 49.
 PARRATT, L. G., 1939, *Phys. Rev.*, **56**, 295.
 RITCHMYER, F. K., BARNES, S. W., and RAMBERG, E., 1934, *Phys. Rev.*, **46**, 843.
 SANDSTRÖM, A., 1935, *Thesis, Upsala*, p. 73.
 STEPHENSON, S. T., 1936, *Phys. Rev.*, **50**, 790.
 VELDKAMP, J., 1935, *Physica*, **2**, 25.
 VIDAL, M., 1949, *Diplome d'Etudes Supérieures, Paris*.

CXVII. Notices of New Books and Periodicals received.

A Textbook on Heat. By J. H. AWBERY. [Pp. 298.] (London: Longmans, 1949.) Price 15s.

IN this refreshingly original treatment the author sets out to emphasize general principles rather than the detailed behaviour of particular substances. The first nine chapters are an excellent and a clear exposition of the underlying physical principles. The usefulness of the book as a whole, however, is somewhat impaired by a very variable level of treatment, ranging from a chapter on thermal expansion at about Higher School Certificate level to a chapter on heat conduction, introducing complex variable methods.

This lack of balance will not prevent the book becoming a "recommended text," but may cause it to appear in the section for "supplementary reading." N. T.

General Psychology. By W. J. H. SPROTT. [Pp. ix+467.] Second Edition. (London : Longmans Green & Co. Ltd., 1947.) Price 14s.

THIS work was first published in 1937, was reprinted in 1942 and now a second edition has been called for. This speaks for itself: the text-book has met a real need.

The second half of the book, which deals principally with perception, sensation and the thought processes, remains unchanged: but the earlier chapters, in which the springs of action are discussed and the dynamics of behaviour are traced, have been modified considerably and in large part re-written. The result is a book which can be recommended as an excellent introductory text-book to any serious student of General Psychology.

N. K. M.

Elements of Aerodynamics of Supersonic Flows. By ANTONIO FERRI. [Pp. 434.] (New York : Macmillan.) Price 50s.

THIS book, written from the standpoint of the engineer, deals with the gas dynamics of flow at speeds greater than sound. For the most part attention is confined to inviscid non-conducting gases, though the problem of the interaction of boundary layer and shock wave is not overlooked.

The early chapters are concerned with the general physical principles underlying high-speed flow phenomena, including a descriptive treatment of the method of characteristics and their application to the solution of specific problems including rotational flows, an account of the theory of shock waves and the use of shock polars and problems arising from the interaction of shock waves. This is followed by a chapter on the methods of measurement. The remainder of the book is devoted to a detailed examination of specific problems, including the behaviour of effusers, diffusers, supersonic wings and projectiles.

L. H.

Calculating Instruments and Machines. By D. R. HARTREE. (Urbana : University of Illinois Press.) Price \$4.50.

THIS book falls into two sections, as indicated by its title, the first dealing with calculating instruments, or analogue machines as they are often called, and the second with digital machines. It will be particularly useful to those who, while not concerned directly with the design and operation of such machines wish some knowledge of their principles and capabilities.

In view of this restricted aim, the historical portions of the book, as those concerned with details of individual machines, are selective rather than comprehensive. In the first portion, attention is given mainly to the differential analyser, and in the second, to the American digital machines. A full account of the principles of operation is followed by examples of the "breaking down" of calculations for the machines and indications from the author's experience of the difficulties met in practical computation.

The eighth chapter, on "Projects and Prospects" contains much interesting information on immediate developments. A flaw in the book is that the attentive reader occasionally wishes to follow a reference and finds it missing from the bibliography. This detracts little from its value.

G. W.

[The Editors do not hold themselves responsible for the views expressed by their correspondents.]

INDEX TO VOL. XL.

- ALEXOPOULOS (K.)** and Scouloudi (H.), scattering of X-rays by the L electrons of boron, 115.
- Algebraic equations, numerical solution (Porter and Mack), 578.
- Alpha particles, short range, from light elements under proton bombardment (Burcham and Freeman), 807.
- Alloys, change of resistance during ageing (Matyáš), 324.
- Andrews (J. P.), *see* Hogarth and Andrews, 273.
- Angus (J.), *see* Curran, Angus and Cockroft, 36, 53.
- , *see* Curran, Cockroft and Angus, 929.
- , Cockroft (A. L.) and Curran (S. C.), investigations of soft radiations by proportional counters. —III. The β spectrum of carbon, 14, 522.
- Ansbacher (F.) and Ehrenberg (W.), derivation of statistical expressions from Gibbs' canonical ensemble, 626.
- Approximate solutions to physical problems (Lighthill), 1179.
- Atmospherics, wave forms. (Khastgir and Roy), 1129.
- Atkinson (F. V.), Sommerfeld's "radiation condition", 645.
- Barium titanate, symmetry changes at low temperatures and their relation to ferroelectric properties (Kay and Vousden), 1019.
- , theory.—Pt. I. (Devonshire), 1040.
- Bellman (R.), Marshak (R. E.) and Wing (G. M.), Laplace transform solution of two-medium neutron ageing problem, 297.
- Beta-rays, focusing in a prolate spheroidal magnetic field (Richardson, H. O. W.), 233.
- Bhatnagar (P. L.) and Singwi (K. S.), distance correlations in an ideal Fermi-Dirac gas, 917.
- Billig (E.) and Plessner (K. W.), efficiency of selenium barrier-photo-cell used as a converter of light into electrical energy, 568.
- Bismuth fibres, magneto resistance and crystalline structure (Donovan and Conn), 283.
- Blackett (P. M. S.), magnetic field of massive rotating bodies, 125.
- Bolton (H. C.), *see* Yarnold and Bolton, 956.
- Bonetti (A.) and Dilworth, Miss (C.), heavy splinters in cosmic ray stars, 585.
- Books, new :—J. W. Archbold, Algebraic geometry of a plane, 245 ; H. D. Anthony, Science and its background, 684 ; Anuario del Observatorio astronomico de Madrid, 782 ; J. H. Awbery, A textbook on heat, 1269 ; E. Bauer, L'Electromagnetisme, hier et aujourd'hui, 782 ; S. Bhagavantum and T. Venkatarayudu, Theory of groups and its applications, 681 ; M. Born, Natural Philosophy of cause and chance, 683 ; E. Castelli, Existentialisme Théologique, 123 ; C. Cherry, Pulses and transients in communication networks, 781 ; L. J. Comrie, Six figure mathematical tables, 588 ; J. L. Coolidge, Mathematics of great amateurs, 782 ; R. Courant and K. O. Friedrichs, Supersonic flow and shock waves, 681 ; Cullick, Fundamentals of electromagnetism, 882 ; W. C. Dampier, History of Science, 479 ; N. E. Dorsey, Freezing of supercooled water, 479 ; A. Ferri, Elements of aerodynamics of supersonic flow, 1270 ; B. Finzi and M. Pastori, Calcolo tensoriale e

- applicazioni, 684; J. W. Gibbs, Collected works, 480; G. F. J. Garlick, Luminescent materials, 1178; D. R. Hartree, Calculating instruments and machines, 1270; G. H. Hardy, Divergent series, 782; L. Hartshorn, Radio-frequency heating, 781; W. Heisenberg, Two lectures, 683; Ionospheric radio propagation, 247; Isotopes, 245; R. W. James, Crystalline state, Vol. II., 246; Journal of Physical Society of Japan, 480; E. Justi, Leitfähigkeit und Leitungsmechanismus fester Stoffe, 882; K. Lonsdale, Crystals and X-rays, 588; W. Magnus and F. Oberhettinger, Formulæ and theorems for the special functions of mathematical physics, 969; L. Marton, Advances in electronics, 588; A. C. Merrington, Vicometry, 781; E. A. Milne, Vectorial mechanics, 246; J. von Neumann, Les fondements mathématique de la mécanique quantique, 679; Progress of physics, Physical Society report, 1178; A. D. Ritchie, Essays in Philosophy, 480; G. S. Rushbrooke, Introduction to statistical mechanics, 968; Science and Technology in China, 479; Scott Blair, G. W. Survey of general and applied rheology, 1093; W. J. H. Sprott, General Psychology, 1270; Surface Chemistry (various authors), 968; Tables of Bessel Functions of fractional order, Vol. I., 124; A. Tarski, Cardinal Algebras, 781; J. T. Toohey, Elementary handbook of Logic, 684; L. R. G. Treloar, Physics of rubber elasticity, 1177; T. Vogel, Vibrations of elastic systems in a sound-field, 247; L. L. Whyte, Unitary principle in physics and biology, 882; E. Whittaker, From Euclid to Eddington, 969; G. Wentzel, Quantum theory of fields, 1177; H. Yukawa, Progress of theoretical physics, 681; G. Zener, Elasticity and an elasticity of metals, 588.
- Boswell (I. I.), *see* Partington, Planer and Boswell, 157,
- Bosworth (R. C. L.), distribution of reaction times for turbulent flow in cylindrical reactors, 314.
- Boulton (T. H.), *see* Hume-Rothery and Boulton, 71.
- Brillouin zones for CO_2Al_3 and NiAl_3 (Raynor and Waldron), 198.
- Brown, Miss (R. H.), Camerini (U.), Fowler (P. H.), Heitler (H.), King (D. T.) and Powell (C. F.), nuclear transmutations produced by cosmic-ray particles of great energy.—Part I. Observations with photographic plates exposed at an altitude of 11,000 feet, 862.
- Burcham (W. E.) and Freeman (Joan M.), emission of short-range α particles from light elements under proton bombardment.—I. Experimental method and the reaction $^{10}\text{B}(p\alpha)^7\text{Be}$, 807.
- Burge (E. J.) and Snellman (O.), magnetic double refraction of liquid mixtures.—I. Benzene in various solvents, 994.
- Cabrera (N.), oxidation of metals at low temperatures and influence of light, 175.
- Caffyn (J. E.), *see* Scott Blair and Caffyn, 80.
- Camerini (U.), *see* Brown *et alia*, 862.
- *et alia*, nuclear transmutations produced by cosmic-ray particles of great energy.—Part II. Observations at high altitudes by means of free balloons, 1073.
- Carbon arc, high intensity crater characteristics (Francis, Hawkins and Willoughby), 546.
- Cauchois (Y.) and Mott (N. F.), interpretation of X-ray absorption spectra of solids, 1260.
- Chirgwin (B. H.), *see* Kilminster and Chirgwin, 226.
- "Clock paradox" of special relativity, so-called (Milne and Whitrow), 1244.
- Coates (A. C.) and Herz (R. H.), experiments with nuclear-track emulsions sensitive at minimum ionizing power, 1088.
- Cockroft (A. L.), *see* Curran, Angus and Cockroft, 36, 53.
- , *see* Angus, Cockroft and Curran, 522.

- Cockroft (A. L.), *see* Curran, Cockroft and Angus, 929.
- and Insch (G. W.), investigation of soft radiations by proportional counters.—VI. The β spectrum of sulphur, 35, 1014.
- Collision problems, use of canonical transformations, (Heitler and Ma), 651.
- Communication theory, note, (Weston), 449.
- Conn (G. K. T.), *see* Donovan and Conn, 283.
- Coor (T.), *see* Camerini *et alia*, 1073.
- Corinaldesi (E.) and Field (G.), scattering of pseudoscalar charged mesons by nucleons.—I., 1159.
- Cosmic rays.
- Angular distribution, frequency and absorption of slow protons (Lattimore), 394.
- Origin of stars (Harding), 530.
- Heavy splinters in stars (Bonetti and Dilworth), 585.
- Production of stars (Thomson, G. P.), 589.
- Nuclear transmutations by particles of great energy.—Part I. (Brown *et alia*), 862. Part II. (Camerini *et alia*), 1073.
- Determination of charge of heavy particles (Sörensen), 947.
- Coulson (C. A.) and Moffit (W. E.), properties of certain strained hydrocarbons, 1.
- and Fischer, Miss (I.), notes on the molecular orbital treatment of the hydrogen molecule, 386.
- and Longuet-Higgins (H. C.), notes on validity and application of the method of molecular orbitals, 1172.
- Cullen (A. L.), channel section waveguide radiator, 417.
- Curran (S. C.), Angus (A.) and Cockroft (A. L.), investigations of soft radiations.—I., 36.
- —II., the beta spectrum of tritium, 53.
- , *see* Angus, Cockroft and Curran, 522.
- , *see* Wilson and Curran, 631.
- Curran (S. C.), Cockroft (A. L.) and Angus (J.), investigation of soft radiation by proportional counters.—V. Use as a detector of ultra-violet quanta and analysis of the gas multiplication process, 929.
- Davies (J. G.) and Ellyett (C. D.), diffraction of radio waves from meteor trails and measurement of meteor velocities, 614.
- Davies (J. H.), *see* Camerini *et alia*, 1073.
- , Lock (W. O.) and Muirhead (H.), decay of μ -mesons, 1250.
- Deas (H. D.) and Emeleus (K. G.), determination of the electron energy distribution in gases from Townsend's ionization coefficient, 460.
- Devonshire (A. F.), theory of barium titanate.—Pt. I., 1040.
- Dielectric and piezoelectric constants in crystals, lattice theory (Huang), 733.
- properties, titanates of perovskite type (Partington, Planer, Boswell), 157.
- Dielectrics, mechanical forces in, (Smith-White), 466.
- Dilworth, Miss (C.), *see* Bonetti and Dilworth), 585.
- Dimensions of physical magnitudes (Dingle), 94.
- Dingle (H.), dimensions of physical magnitudes, (a paradox in dimensional theory), 94.
- Dingle (R. B.), zero-point energy of a system of particles, 573.
- Discharges in high pressure mercury and cadmium vapour, theory, (Francis), 435.
- Discharge, high pressure helium, theory (Francis), 1063.
- Dislocations in anisotropic materials, edge (Eshelby), 903.
- Donovan (B.) and Conn (G. K. T.), electrical conductivity of bismuth fibres.—I. Magneto-resistance and the crystalline structure, 283.
- Dudley (B. R.) and Swift (H. W.), frictional relaxation oscillations, 849.
- Ehrenberg (W.), *see* Ansbacher and Ehrenberg, 626.

- Electrodynamics in a rotating frame of reference (Trocheris), 1143.
- Electron energies in highly ionized gas (Giovannelli), 206.
- in gases (Deans and Emeleus), 460.
- Ellyett (C. D.), *see* Davies and Ellyett, 614.
- Emeleus (K. G.), *see* Deas and Emeleus, 460.
- Eshelby (J. D.), edge dislocations in anisotropic materials, 903.
- Ferromagnetic resonance, theory (Polder), 99.
- Field (G.), *see* Corinaldesi and Field, 1159.
- Field theory, minimum integrals in (Kilminster and Chirgwin), 226.
- Fischer, Miss (I.), *see* Coulson and Fischer, 386.
- Flow behind a stationary shock (Lighthill), 214.
- Following system, optimum parameters (Mack), 922.
- Fowler (P. H.), *see* Brown *et alia*, 862.
- , *see* Camerini *et alia*, 1073.
- Francis (V. G.), theory of the high pressure mercury vapour and cadmium vapour discharges, 435.
- , theory of the high pressure helium discharge, 1063.
- , Hawkins (F. S.) and Willoughby (A. H.), characteristics of the crater of the high intensity carbon arc, 546.
- Freeman (Joan M.), *see* Burcham and Freeman, 807.
- Frictional relaxation oscillations (Dudley and Swift), 849.
- Fürth (R.), theory of strength of quasi-isotropic solids, 1227.
- Galt (J. K.), note on the effect of impurities and cold work on the thermoelectric power of aluminium, 309.
- Giovannelli (R. G.), electron energies resulting from an electric field in a highly ionized gas, 206.
- Hammersley (J. M.), numerical reduction of non-singular matrix pencils, 783.
- Hardie (A. M.), transient response of an oscillatory circuit with re-current discharge, 748.
- Harding (J. B.), origin of cosmic ray stars, 530.
- Hawkins, (Miss D. B. G.), an improved Schmidt plate, 670.
- Hawkins (F. S.), *see* Francis, Hawkins and Willoughby, 546.
- Heitler (H.), *see* Brown *et alia*, 862.
- Heitler (W.) and Ma (S. T.), use of canonical transformations for collision problems, 651.
- Helium II., theory (Mott), 61.
- Herz (R. H.), *see* Coates and Herz, 1088.
- Hill (R.), plastic distortion of non-uniform sheets, 971.
- Hogarth (C. A.) and Andrews (J. P.), variation with oxygen pressure of the thermoelectric power of cadmium oxide, 273.
- Huang (K.), lattice theory of dielectric and piezoelectric constants in crystals, 733.
- Huby (R.), capture of negative mesons, 685.
- Hume-Rothery (W.) and Boulton (T. H.), coefficients of expansion of some solid solutions in aluminium, 71.
- Ideal Fermi-Dirac gas, distance correlations (Bhatnagar and Singwi), 917.
- Insch (G. M.), *see* Cockroft and Insch, 1014.
- Interchange interaction and collective electron ferromagnetism (Wohlfarth), 703.
- Ionization and charge exchange by fast ions of hydrogen and helium (Keene), 369.
- Kay (H. F.) and Vouseh (P.), symmetry changes in barium titanate at low temperatures and their relation to its ferroelectric properties, 1019.
- Keene (J. P.), ionization and charge exchange by fast ions of hydrogen and helium, 369.
- Khastgir (S. R.) and Roy (R.), study of wave forms of atmospherics, 1129.
- Kilminster (C. W.) and Chirgwin (B. H.), note on minimum integrals in field theory, 226.
- King (D. T.), *see* Brown *et alia*, 862.

- Lattimore (S.), angular distribution, frequency and absorption of slow single cosmic ray protons, 394.
- Le Couteur (K. J.) and Rosenfeld (L.), canonical transformations of the Hamiltonian in meson field theory, 151.
- Lighthill (M. J.), flow behind a stationary shock, 214.
- , thermal stresses in turbine blades, 770.
- , technique for rendering approximate solutions to physical problems uniformly valid, 1179.
- , shock strength of supersonic "conical fields", 1202.
- Lock (W. O.), *see* Camerini *et alia*, 1073.
- , *see* Davies, Lock and Muirhead, 1250.
- Longuet-Higgins (H. C.), *see* Coulson and Longuet-Higgins, 1172.
- Ma (S. T.), *see* Heitler and Ma, 651.
- , vacuum polarization in positron theory, 1112.
- MacDonald (D. K. C.), transit-time deterioration of space-charge reduction of shot effect, 561.
- Macfarlane (G. G.), application of Mellin transforms to the summation of slowly convergent series, 188.
- Mack (C.), *see* Porter and Mack, 578.
- , calculation of optimum parameters for a following system, 922.
- M'Ewen (E.), stresses in elastic cylinders in contact along a generatrix, 454.
- Magnetic double refraction in non-polar liquids, theory (Snellman), 983.
- double refraction of liquid mixtures.—I. Benzene in various solvents (Burge and Snellman), 994.
- —II. Toluene, carbon bisulphide, relation to refractive index, 1233.
- field of massive rotating bodies (Blackett), 125.
- problems of axial symmetry (Peierls and Skyrme), 269.
- properties of Ni-Co and related alloys (Wohlfarth), 1095.
- Magneto-thermal effects in ferromagnetics (Stones and Rhodes), 481.
- Marshak (R. E.), *see* Bellman, Marshak and Wing, 297.
- Matyáš (Z.), change of electrical resistance of alloys during ageing, 324.
- Meltzer (B.), note on the identity of thermal noise and shot noise, 1224.
- Mesons.
- Meson field theory (Le Couteur and Rosenfeld), 151.
- Mechanism of π -meson disintegrations (Perkins), 601.
- Motion of vector meson in magnetic field (Symonds), 636.
- Capture of negative mesons (Huby), 685.
- Interaction between mesons (Tzu), 717.
- Loss of energy of slow negative mesons in matter (Rosenberg), 759.
- Scattering of pseudoscalar mesons by nucleons (Corinaldesi and Field), 1159.
- Decay of μ mesons (Davies, Lock and Muirhead, 1250.
- Meteor trails, diffraction of radio waves from (Davies and Ellyett), 614.
- Michell rectangular and sector-shaped pads, Reynolds equation for (Wood), 220.
- Milne (E. A.) and Whitrow (G. J.), the so-called "Clock paradox" of special relativity, 1244.
- Mitchell (J. W.), properties of silver halides containing traces of silver sulphide, 249.
- , lattice defects in silver halide crystals, 667.
- Mitchell (K.), tables of $\int_0^z \frac{-\log(1-y)}{y} dy$ and the properties of this and related functions, 351.
- Moffit (W. E.), *see* Coulson and Moffit, 1.
- Mott (N. F.), a contribution to the theory of helium II, 61.
- , *see* Cauchois and Mott, 1260.
- Molecular orbital treatment of hydrogen molecule (Coulson and Fisher), 386.
- Molecular orbitals, validity and application of the method (Coulson and Longuet-Higgins), 1172.

- Muirhead (H.), *see* Camerini *et alia*, 1073.
 —, *see* Davies, Lock and Muirhead, 1250.
- Non-singular matrix pencils, numerical reduction (Hammersley), 783.
- Non-symmetric stress-energy-momentum tensor and spin-density (Papapetrou), 937.
- Notched plate of an aelotropic material (Okubo), 913.
- Nuclear-track emulsions sensitive at minimum ionizing powder (Coates and Herz), 1088.
- Okubo (H.), notched plate of an aelotropic material, 913.
- Oxidation of metals at low temperatures (Cabrera), 175.
- Pair creation, internal, in Na^{24} (Rae), 1155.
- Papapetrou (A.), non-symmetric stress-energy-momentum tensor and spin-density, 937.
- Partington (J. R.), Planer (G. V.) and Boswell (I. I.), anomalous dielectric properties of polycrystalline titanates of perovskite type, 157.
- Peierls (R. E.) and Skyrme (T. H. R.), tank model for magnetic problems of axial symmetry, 269.
- Perkins (D. H.), mechanism of π -meson disintegrations, 601.
- Planer (G. V.), *see* Partington, Planer and Boswell, 157.
- Plastic distortion of non-uniform sheets (Hill), 971.
- Plessner (K. W.), *see* Billig and Plessner, 568.
- Polder (D.), on the theory of ferromagnetic resonance, 99.
- Porter (A.) and Mack (C.), new methods for numerical solution of algebraic equations, 578.
- Positron theory, vacuum polarization (Ma), 1112.
- Powell (C. F.), *see* Brown *et alia*, 862.
- Quasi-properties, application of theory to anomalous strain-stress relations (Scott Blair and Caffyn), 80.
- Radiation emitted by electron in a parabolic orbit, Kramer's formula (Westfold), 698.
- Radio-frequency radiation in ionized gas with applications to solar atmosphere (Smerd and Westfold), 831.
- Rae (E. R.), internal pair creation in Na^{24} , 1155.
- Raynor (G. V.) and Waldron (M. B.), Brillouin zones for the CO_2Al_9 and NiAl_3 structures, 198.
- , and Wakeman (D. W.), primary solid solution of silver in aluminium, 404.
- Reaction times for turbulent flow in cylindrical reactors (Bosworth), 314.
- Rhodes (P.), *see* Stoner and Rhodes, 481.
- Richardson (H. O. W.), focusing of β -rays in a prolate spheroidal magnetic field, 233.
- Roscoe (R.), flow of viscous fluids round plane obstacles, 338.
- Rosenberg (R. L.), loss of energy of slow negative mesons in matter, 759.
- Rosenfeld (L.), *see* Le Couteur and Rosenfeld, 151.
- Roy (R.), *see* Khastgir and Roy, 1129.
- Scattering of X-rays by L electrons of boron (Alexopoulos and Scouloudi), 115.
- Schmidt plate, improved, (Hawkins), 670.
- Scott Blair (G. W.) and Caffyn (J. E.), application of theory of quasi-properties to treatment of anomalous strain-stress relations, 80.
- Scouloudi (H.), *see* Alexopoulos and Scouloudi, 115.
- Selenium barrier-photocell efficiency (Billig and Plessner), 568.
- Series, summation of slowly convergent (Macfarlane), 188.
- Shock strength of supersonic conical fields (Lighthill), 1202.
- Silver halides, properties (Mitchell J. W.), 249; lattice defects (Mitchell, J. W.), 667.
- Singwi (K. S.), *see* Bhatnagar and Singwi, 917.
- Skyrme (T. H. R.), *see* Peierls and Skyrme, 269.
- Slow-positive ion beams, energy distribution (Yarnold and Bolton), 956.

- Smerd (S. F.) and Westfold (K. C.), characteristics of radio-frequency radiation in an ionized gas, with applications to the transfer of radiation in the solar atmosphere, 831.
- Smith-White (W. B.), on the mechanical forces in dielectrics, 466.
- Snellman (O.), theory of magnetic double refraction in non-polar liquids, 983.
- , *see* Burge and Snellman, 994, 1233.
- Soft radiation, investigations by proportional counters.—I. (Curran, Angus and Cockfort), 36; II. Beta spectrum tritium (Curran, Angus and Cockfort), 53; III. Beta spectrum of C 14 (Angus, Cockfort and Curran), 522; IV. Beta spectrum of Ni 63 (Wilson and Curran), 63; V. Use as detector of U-V. quanta and analysis of multiplication process (Curran, Cockfort and Angus), 929; VI. Beta spectrum of S 35 (Cockfort and Insch), 1014.
- Solid solutions in aluminium, coefficients of expansion (Hume-Rothery and Boulthbee), 71.
- solution of silver in aluminium, the primary (Raynor and Wakeman), 404.
- Sommerfeld's "radiation condition" (Atkinson), 645.
- Sörensen (S. O. C.), determination of the charge of heavy particles emitted during the explosive disintegration of nuclei, 947.
- Space-times of general relativity, regraduation in spherically symmetric (Willmore), 428.
- Statistical expressions from Gibbs' canonical ensemble (Ansbacher and Ehrenberg), 626.
- Stoner (E. C.) and Rhodes (P.), magneto-thermal effects in ferromagnetics, 481.
- Strained hydrocarbons, properties of (Coulson and Moffitt), 1.
- Strength of quasi-isotropic solids, theory (Fürth), 1227.
- Stresses in elastic cylinders (M'Ewen), 454.
- and strains in tube-drawing (Swift), 883.
- Swift (H. W.), stresses and strains in tube-drawing, 883.
- , *see* Dudley and Swift, 849.
- Symonds (N.), motion of a vector meson in a homogeneous magnetic field, 636.
- Tables and properties of $\int_0^2 \frac{-\log(1-y)}{y} dy$ (Mitchell, K.), 351.
- Thermal noise and shot noise (Meltzer), 1224.
- stresses in turbine blades (Lighthill), 770.
- Thermoelectric power of aluminium effect of impurities and cold work (Galt), 309.
- power of cadmium oxide, variation with oxygen pressure (Hogarth and Andrews), 273.
- Thomson (Sir G.), production of cosmic ray stars, 589.
- Tobin (N.), *see* Camerini *et alia*, 1073.
- Transient response of oscillatory circuit with recurrent discharge (Hardie), 748.
- Transit-time deterioration of space-charge reduction of shot effect (MacDonald), 561.
- Trocheris (M. G.), electrodynamics in a rotating frame of reference, 1143.
- Two-medium neutron ageing problem (Bellman, Marshak and Wing), 297.
- Tzu (H. Y.), interaction between mesons, 717.
- Viscous fluids, flow round plane obstacles (Roscoe), 338.
- Vousden (P.), *see* Kay and Vousden, 1019.
- Wakeman (D. W.), *see* Raynor and Wakeman, 404.
- Waldron (M. B.), *see* Raynor and Waldron, 198.
- Wave guide radiator, channel section, (Cullen), 417.
- Westfold (K. C.), Kramer's formula for the frequency distribution of radiation emitted by an electron in parabolic orbit, 698.
- , *see* Smerd and Westfold, 831.
- Weston (J. D.), a note on the theory of communication, 449.
- Whitrow (G. J.), *see* Milne and Whitrow, 1244.

- Wing (G. M.), *see* Bellman, Marshak and Wing, 297.
- Willmore (T. J.), regraduation in spherically symmetric space-times of general relativity, 428.
- Willoughby (A. H.), *see* Francis, Hawkins and Willoughby, 546.
- Wilson (H. W.) and Curran (S. C.), investigations of soft radiations by proportional counter.—IV. The β spectrum of Ni^{63} , 631.
- Wohlfarth (E. R.), interchange interaction collective electron ferromagnetism, 703.
- Wohlfarth (E. R.), magnetic properties of Nickel-Cobalt and Related alloys, 1095.
- Wood (Mrs. W. L.), note on a new form of the solution of Reynolds' equation for Michell rectangular and sector-shaped pads, 220.
- X-ray absorption spectra of solid, Interpretation (Cauchois and Mott) 1260.
- Yarnold (G. D.) and Bolton (H. C.), energy distribution of slow-positive ion beams, 956.
- Zero-point energy of system of particles (Dingle), 573.

END OF THE FOURTEENTH VOLUME.